

# Bayesian Nonparametric Estimation of a Unimodal Density via two $\mathbf{S}$ -paths<sup>1</sup>

MAN-WAI HO

(February 2, 2008)

## Abstract

A Bayesian nonparametric method for unimodal densities on the real line is provided by considering a class of species sampling mixture models containing random densities that are unimodal and not necessarily symmetric. This class of densities generalize the model considered by Brunner (1992), in which the Dirichlet process is replaced by a more general class of species sampling models. A novel and explicit characterization of the posterior distribution via a finite mixture of two dependent  $\mathbf{S}$ -paths is derived. This results in a closed-form and tractable Bayes estimator for any unimodal density in terms of a finite sum over two  $\mathbf{S}$ -paths. To approximate this class of estimates, we propose a sequential importance sampling algorithm that exploits the idea of the accelerated path sampler, an efficient path-sampling Markov chain Monte Carlo method. Numerical simulations are given to demonstrate the practicality and the effectiveness of our methodology.

A Bayesian nonparametric method for unimodal densities on the real line is provided by considering a class of species sampling mixture models containing random densities that are unimodal and not necessarily symmetric. This class of densities generalize the model considered by Brunner (1992), in which the Dirichlet process is replaced by a more general class of species sampling models. A novel and explicit characterization of the posterior distribution via a finite mixture of two dependent

---

<sup>1</sup>AMS 2000 subject classifications. Primary 62G05; secondary 62F15.

*Key words and phrases.* Species sampling model, Species sampling mixture model, Rao–Blackwellization, Sequential importance sampling, Markov chain Monte Carlo, Accelerated path sampler.

**S**-paths is derived. This results in a closed-form and tractable Bayes estimator for any unimodal density in terms of a finite sum over two **S**-paths. To approximate this class of estimates, we propose a sequential importance sampling algorithm that exploits the idea of the accelerated path sampler, an efficient path-sampling Markov chain Monte Carlo method. Numerical simulations are given to demonstrate the practicality and the effectiveness of our methodology.

## 1 Introduction

Statistical theory usually assumes that data come from a distribution that is symmetric and unimodal at zero, such as a normal distribution or a Student's  $t$  distribution. However, it is common in real-life applications that underlying distribution of response variable, even though unimodal, may not be symmetric about its mode which is different from zero. For more information, see Dharmadhikari and Joag-Dev (1988) and Bertin, Cuculescu and Theodorescu (1997). There is a vast amount of literature on nonparametric estimations of unimodal densities and the mode from a frequentist viewpoint including early works of Granander (1956), Parzen (1962), Chernoff (1964), Robertson (1967), Venter (1967), Prakasa Rao (1969), Wegman (1969, 1970a, 1970b, 1971), other further studies by Lye and Martin (1993), Bickel and Fan (1996), Wang (1996) and Birgé (1997) and among others. Some recent methods are, for example, a recursive method in Cheng, Gasser, and Hall (1999), kernel-based methods in Hall and Huang (2001, 2002), and other parametric models in Fernández and Steel (1998), Jones (2004) and Ferreira and Steel (2006).

From a Bayesian viewpoint, Brunner (1992) gave a nonparametric solution to the problems by assuming a mixture representation same as that in (1) wherein the mixing distribution  $G$  is a Dirichlet process (Ferguson (1973)) for a unimodal density with a general mode  $\theta$  on the real line  $\mathcal{R}$ . The posterior distribution and the Bayes estimate of the unimodal density can be characterized in terms of random partitions (see, e.g, Lo (1984) and Lo, Brunner and Chan (1996) for these well-

established theoretical results on Dirichlet process mixture models).

This paper is devoted to improving Brunner's results and developing an efficient numerical method for practical usage of the Bayes solutions. A class of unimodal densities with mode at  $\theta$  of interest is defined by

$$f(t|G, \theta) = \int \frac{1}{X} [\mathbb{I}(0 < t - \theta \leq X) - \mathbb{I}(X \leq t - \theta < 0)] G(dX), \quad t \in \mathcal{R}, \quad (1)$$

where  $\mathbb{I}(B)$  is the indicator of an event  $B$  and  $G$  is from the class of species sampling models developed in Pitman (1995, 1996), of which the Dirichlet process is a member. All the results follow are therefore applicable to Brunner's model as his model is a special case of (1). The validity of the mixture representation for all unimodal densities given in the right hand of (1) can be justified by noting equality between its integral when  $\theta = 0$  and the distribution function of any unimodal density with mode at zero given in Feller (1971, page 158).

The posterior distribution of (1), like Brunner's model, can also be characterized in terms of random partitions, as the models are special cases of the species sampling mixture model defined in Ishwaran and James (2003) which takes the same form as (1) with the kernel  $X^{-1} [\mathbb{I}(0 < t - \theta \leq X) - \mathbb{I}(X \leq t - \theta < 0)]$  replaced by any density function in  $t$  given  $\theta$  and  $X$ . In this work, by utilizing the special and nice features of the kernel in (1) (see (7)) and noticing irrelevancy of some information carried by a partition in characterizing the posterior distribution, we are able to refine the partition-based results to show that the unimodal densities possess special structures related to two **S**-paths, where an **S**-path is a random vector defined in Brunner and Lo (1989) (see also Dykstra and Laud (1981). Generally speaking, there exists a tractable characterization of the posterior distribution via some combinatorial structures that are considerably less complex than partitions. Such a characterization is known to be the first explicit type that is based on two **S**-paths. Similar phenomena based on one single **S**-path could be found in Bayes estimations of symmetric unimodal or decreasing densities by Brunner and Lo (1989), Brunner (1995) and Ho (2006b) and monotone hazard functions by Dykstra and

Laud (1981), Lo and Weng (1989) and Ho (2006a), as the space of  $\mathbf{S}$ -paths is considerably smaller than that of partitions (Brunner and Lo 1989). Intuitively, this characterization depending on two  $\mathbf{S}$ -paths can be explained by the fact that there are two (possibly different) non-increasing curves on each side of the mode in unimodal densities, but not only one (identical on either sides) in symmetric unimodal densities of which can be characterized in terms of one  $\mathbf{S}$ -path (Albert Y. Lo, private conversation).

It is recognized that if one could efficiently sample the two  $\mathbf{S}$ -paths in this context, this would lead to more parsimonious methods for inference. Motivated by the co-existence of and the resemblance in constructions of an SIS algorithm and a Gibbs sampler for sampling random partitions in many Bayesian mixture models (Lo, Brunner and Chan (1996) and Ishwaran and James (2003)), we propose (in Section 3) a novel sequential importance sampling (SIS) method (Kong, Liu and Wong (1994) and Liu and Chen (1998)), dubbed *sequential importance path (SIP) sampler*, for sampling directly one single  $\mathbf{S}$ -path in the aforementioned models under monotonicity constraints by borrowing the idea behind the success of an efficient Markov chain Monte Carlo (MCMC) method introduced in Ho (2002, 2006a, 2006b) that serves the same purpose. Then, a natural SIS scheme based on applications of the SIP sampler is introduced for sampling the unknown mode  $\theta$  and the two  $\mathbf{S}$ -paths in evaluating/approximating posterior quantities for models in (1).

### 1.1 Some backgrounds on species sampling models

Pitman (1995, 1996) developed the class of species sampling models that corresponds to the set of all random probability measure of the form

$$P(\cdot) = \sum_k W_k \delta_{V_k}(\cdot) + \left(1 - \sum_k W_k\right) H(\cdot), \quad (2)$$

where  $0 < W_k < 1$  are random weights such that  $\sum_k W_k \leq 1$ , independently of  $V_k$ , which are i.i.d. random variables with some non-atomic distribution  $H$  over a

measurable Polish space, and  $\delta_{V_k}(\cdot)$  is a Dirac probability measure degenerate at  $V_k$ . This class includes a large number of well-known random processes, for instance, the Dirichlet process and its two-parameter extension, called the two-parameter Poisson-Dirichlet process (Pitman and Yor (1997)), the class of finite-dimensional Dirichlet priors discussed in detail in Ishwaran and Zarepour (2002a, 2002b), and the homogeneous normalized random measures with independent increments discussed in Regazzini, Lijoi and Prünster (2003).

Suppose  $\mathbf{X} = (X_1, \dots, X_N)$  is a random sample from (2). The joint marginal distribution of  $\mathbf{X}$  is determined by the prediction rule,  $\Pr\{X_1 \in \cdot\} = H(\cdot)$  and

$$\Pr\{X_{k+1} \in \cdot | X_1, \dots, X_k\} = \ell_{0,k} H(\cdot) + \sum_{j=1}^{N_k} \ell_{j,k} \delta_{X_j^*}(\cdot), \quad k = 2, \dots, N-1, \quad (3)$$

where  $H$  is non-atomic and  $\ell_{0,k}$  and  $\ell_{j,k}$  are non-negative measurable functions of  $X_1, \dots, X_k$ . The above prediction rule conveys that given  $X_1, \dots, X_k$ , which correspond to  $N_k$  unique values  $X_1^*, \dots, X_{N_k}^*$  of respective numbers of duplicates  $e_1, \dots, e_{N_k}$ , then the next observation  $X_{k+1}$  takes the same value as  $X_j^*$  with probability  $\ell_{j,k}$ ,  $j = 1, \dots, N_k$ ; otherwise it takes a new value from  $H$  with probability  $\ell_{0,k}$ . As a consequence of the exchangeability of  $(X_1, \dots, X_N)$ , Pitman (1996) shows that the distribution of  $X_1, \dots, X_N$ , denoted by  $\mu(d\mathbf{X})$ , is uniquely characterized by the joint law of its unique values and an exchangeable partition probability function (EPPF)

$$\pi(\mathbf{p}) = \chi(e_1, \dots, e_{N(\mathbf{p})}) \quad (4)$$

induced by the unique values. That is,

$$\mu(d\mathbf{X}) = \pi(\mathbf{p}) \prod_{k=1}^{N(\mathbf{p})} H(dX_k^*),$$

where  $\mathbf{p} = \{C_1, \dots, C_{N(\mathbf{p})}\}$  of the integers  $\{1, \dots, N\}$  is a partition of  $N(\mathbf{p})$  cells induced by  $C_k = \{j : X_j = X_k^*\}$  and  $\chi$  is a unique symmetric function depending only upon  $e_k$ , the number of elements in or the size of  $C_k$ ,  $k = 1, \dots, N(\mathbf{p})$  (see Pitman (1996) and Ishwaran and James (2003, Section 2) for more information).

## 2 A posterior distribution via S-paths

This section concerns Bayes estimation of a unimodal density on the line  $\mathcal{R}$  with a general mode  $\theta$ , defined by the species sampling mixture model in (1). Suppose we observe  $N$  i.i.d. observations  $\mathbf{T} = (T_1, \dots, T_N)$  from (1) and assume any prior  $\pi(d\theta)$  for  $\theta$ . Given  $\mathbf{T}$ , denote  $\mathcal{P}(dG|\theta, \mathbf{T})$  and  $\mathcal{P}(d\theta|\mathbf{T})$  as the posterior distribution of  $G$  given  $\theta$  and the posterior distribution of  $\theta$ , respectively. The posterior distribution of the pair  $(G, \theta)$  in (1) can always be determined by the double expectation formula,

$$\mathbb{E}[h(G, \theta)|\mathbf{T}] = \mathbb{E}\{\mathbb{E}[h(G, \theta)|\theta, \mathbf{T}|\mathbf{T}]\} = \int_{\mathcal{R}} \int_{\mathcal{M}} h(G, \theta) \mathcal{P}(dG|\theta, \mathbf{T}) \mathcal{P}(d\theta|\mathbf{T}), \quad (5)$$

where  $h$  is any nonnegative or integrable function and  $\mathcal{M}$  is the space of probability measures over  $\mathcal{R}$ . Let us first look at  $\mathcal{P}(dG|\theta, \mathbf{T})$  and then discuss  $\mathcal{P}(d\theta|\mathbf{T})$  later on.

Suppose  $\theta$  is given. We can always assume that

$$(T_1 - \theta, \dots, T_N - \theta) = \mathbf{Z} \cup \mathbf{Y} = (Z_{N-n}, Z_{N-n-1}, \dots, Z_1) \cup (Y_1, Y_2, \dots, Y_n), \quad (6)$$

where  $Z_{N-n} < Z_{N-n-1} < \dots < Z_1 < 0$  and  $0 < Y_1 < Y_2 < \dots < Y_n$ . Denote the missing variables in (1) by  $\mathbf{X} = (X_1, \dots, X_N)$ . It is worthy of note that once an observation is taken from (1), the kernel can be well-simplified according to two mutually exclusive situations, that is, the likelihood of an observation  $T_k$  in  $\mathbf{T}$  is given by

$$f(T_k|G, \theta) = \begin{cases} \int (-X^{-1}) \mathbb{I}(X \leq T_k - \theta) G(dX) & T_k - \theta < 0 \\ \int X^{-1} \mathbb{I}(T_k - \theta \leq X) G(dX) & T_k - \theta > 0. \end{cases} \quad (7)$$

The distinctiveness of the kernel yields a similar simplification (see (33) and (34)) in the posterior distribution of  $G$  given  $\theta$  in terms of partitions  $\mathbf{p}$  of the integers  $\{1, \dots, N\}$  in (32), readily available from Theorems 1 and 2 in Ishwaran and James (2003). This implies that the  $n$  resulting positive observations after subtraction of  $\theta$ ,  $Y_1, \dots, Y_n$ , can only “cluster” with one another but not any negative observation or vice versa. Hence, it is eligible to “split” the partition  $\mathbf{p}$  of the  $N$  integers/observations into two non-overlapping partitions  $\mathbf{p}^+$  and  $\mathbf{p}^-$ . Write

$\mathbf{p} = \mathbf{p}^+ \cup \mathbf{p}^-$ . Without loss of generality, assume that  $\mathbf{p}^+ = \{C_1, \dots, C_{N(\mathbf{p}^+)}\}$  denotes the partition of the  $n$  positive observations and  $\mathbf{p}^- = \{C_{N(\mathbf{p}^+)+1}, \dots, C_{N(\mathbf{p})}\}$  of the remaining  $N - n$  negative observations. Define

$$\pi(\mathbf{p}^- | \mathbf{p}^+) := \frac{\chi(e_1, \dots, e_{N(\mathbf{p}^+)}, e_{N(\mathbf{p}^+)+1}, \dots, e_{N(\mathbf{p})})}{\chi(e_1, \dots, e_{N(\mathbf{p}^+)})} = \frac{\pi(\mathbf{p})}{\pi(\mathbf{p}^+)}, \quad (8)$$

where  $\pi(\cdot)$  is defined in (4), such that

$$\pi(\mathbf{p}) = \pi(\mathbf{p}^- | \mathbf{p}^+) \times \pi(\mathbf{p}^+). \quad (9)$$

These, together with the facts that the second line of (7) resembles, while the other line is symmetrical to, the scaled mixture of uniform representation of a symmetric unimodal density with mode at zero due to Khintchine (1938) and Shepp (1962), yield a posterior distribution of  $G$  given  $\theta$ , which is expressible in terms of two dependent  $\mathbf{S}$ -paths, as a consequence of applications of Theorem 2.1 and Corollary 2.2 in Ho (2006b).

Let us fix some notation before stating the main results. Define an integer-valued vector  $\mathbf{S} = (S_0, S_1, \dots, S_{n-1}, S_n)$  (Dykstra and Laud (1981) and Brunner and Lo (1989)), referred to as an  $\mathbf{S}$ -path (of  $n + 1$  coordinates), which satisfies (i)  $S_0 = 0$  and  $S_n = n$ ; (ii)  $S_j \leq j$ ,  $j = 1, \dots, n - 1$ ; and (iii)  $S_j \leq S_{j+1}$ ,  $j = 1, \dots, n - 1$ . A path  $\mathbf{S}$  is said to *correspond to* one or many partitions  $\mathbf{p}$  of the integers  $\{1, \dots, n\}$ , provided that (i) labels of the maximal elements of the  $N(\mathbf{p})$  cells in  $\mathbf{p}$  coincide with locations  $j$  at which  $S_j > S_{j-1}$ , and (ii) size  $e_k$  of the cell  $C_k$  for all  $k = 1, \dots, N(\mathbf{p})$  with a maximal element  $j$ ,  $j = 1, \dots, n$ , is identical to  $S_j - S_{j-1}$ . Let  $\mathbb{C}_{\mathbf{S}}$  denote the collection of partitions that correspond to a given  $\mathbf{S}$ . Then, the total number of partitions in  $\mathbb{C}_{\mathbf{S}}$  is given by (Brunner and Lo (1989))

$$|\mathbb{C}_{\mathbf{S}}| = \prod_{j=1: S_j > S_{j-1}}^n \binom{j-1-S_{j-1}}{j-S_j}. \quad (10)$$

See Ho (2002) for more discussion of the relation between  $\mathbf{p}$  and  $\mathbf{S}$ . Following from the symmetric definition of  $\chi$  in (4), we have

$$\pi(\mathbf{p}) = \chi(e_1, \dots, e_{N(\mathbf{p})}) := \chi(\mathcal{M}_{1,n}(\mathbf{S})), \quad \text{if } \mathbf{p} \in \mathbb{C}_{\mathbf{S}}, \quad (11)$$

where, for any integer  $1 \leq a < b \leq n$ ,

$$\mathcal{M}_{a,b}(\mathbf{S}) = \{S_j - S_{j-1} : S_j > S_{j-1}, j = a, a+1, \dots, b\}.$$

Write  $\sum_{\mathbf{S}}$  as summing over all paths  $\mathbf{S}$  of the same number of coordinates, and  $\prod_{\{j^*|\mathbf{S}\}}$  and  $\sum_{\{j^*|\mathbf{S}\}}$  as  $\prod_{j=1:n: S_j > S_{j-1}}^n$  and  $\sum_{j=1:n: S_j > S_{j-1}}^n$  conditioning on  $\mathbf{S}$ , respectively.

**Theorem 2.1.** Suppose  $\theta$  is given and  $\mathbf{T}$  are  $N$  i.i.d. observations from (1). That is, (6) holds. Then, the distribution of  $\mathbf{X}$  given  $\theta$  and  $\mathbf{T}$  can be summarized by a joint law of  $(\mathbf{V}, \mathbf{U}), (\mathbf{S}^-, \mathbf{S}^+)|\theta, \mathbf{T}$  defined as follows.

- (i) Given  $(\theta, \mathbf{T})$ , two paths  $\mathbf{S}^+ = (0, S_1^+, \dots, S_{n-1}^+, n)$  and  $\mathbf{S}^- = (0, S_1^-, \dots, S_{N-n-1}^-, N-n)$  of  $n+1$  and  $N-n+1$  coordinates, respectively, have a (discrete) joint distribution  $\pi(\mathbf{S}^-, \mathbf{S}^+|\theta, \mathbf{T}) \propto \phi_\theta^+(\mathbf{S}^+, \mathbf{T}) \times \phi_\theta^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T})$ , where

$$\phi_\theta^+(\mathbf{S}^+, \mathbf{T}) = |\mathbb{C}_{\mathbf{S}^+}| \chi(\mathcal{M}_{1,n}(\mathbf{S}^+)) \prod_{\{j^*|\mathbf{S}^+\}} \int_{Y_j}^\infty U_j^{-(S_j^+ - S_{j-1}^+)} H(dU_j) \quad (12)$$

and

$$\phi_\theta^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T}) = |\mathbb{C}_{\mathbf{S}^-}| \frac{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))}{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+))} \prod_{\{j^*|\mathbf{S}^-\}} \int_{-\infty}^{Z_j} (-V_j)^{-(S_j^- - S_{j-1}^-)} H(dV_j) \quad (13)$$

with  $|\mathbb{C}_{\mathbf{S}}|$  and  $\chi(\cdot)$  defined in (10) and (11), respectively.

- (ii) Given  $(\mathbf{S}^-, \mathbf{S}^+)$  and  $(\theta, \mathbf{T})$ , there exist  $N(\mathbf{S}^+) = \sum_{j=1}^n \mathbb{I}(S_j^+ > S_{j-1}^+)$  positive and  $N(\mathbf{S}^-) = \sum_{j=1}^{N-n} \mathbb{I}(S_j^- > S_{j-1}^-)$  negative unique values on  $\mathcal{R}$  among  $\{X_1, \dots, X_N\}$ , denoted by  $\mathbf{U} = \{U_j : S_j^+ > S_{j-1}^+, j = 1, \dots, n\}$  and  $\mathbf{V} = \{V_j : S_j^- > S_{j-1}^-, j = 1, \dots, N-n\}$ , respectively. They are distributed, conditionally independent of one another, as

$$H_j^+(dU_j|\mathbf{S}^+, \mathbf{Y}) \propto \mathbb{I}(Y_j \leq U_j) U_j^{-(S_j^+ - S_{j-1}^+)} H(dU_j), \quad (14)$$

and

$$H_j^-(dV_j|\mathbf{S}^-, \mathbf{Z}) \propto \mathbb{I}(V_j \leq Z_j) (-V_j)^{-(S_j^- - S_{j-1}^-)} H(dV_j), \quad (15)$$

respectively.



**Theorem 2.2.** For any nonnegative or integrable function  $g$ , the law of  $G$  given  $\theta$  and  $\mathbf{T}$  is characterized by

$$\begin{aligned} & \int_{\mathcal{M}} g(G) \mathcal{P}(dG|\theta, \mathbf{T}) \\ &= \sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} \left[ \int_{\mathcal{R}^{N(\mathbf{S}^+) + N(\mathbf{S}^-)}} \left\{ \int_{\mathcal{M}} g(G) \mathcal{P}(dG|\mathbf{V}, \mathbf{U}, \mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T}) \right\} \right. \\ & \quad \left. \prod_{\{j^*|\mathbf{S}^-\}} H_j^-(dV_j|\mathbf{S}^-, \mathbf{Z}) \prod_{\{j^*|\mathbf{S}^+\}} H_j^+(dU_j|\mathbf{S}^+, \mathbf{Y}) \right] \pi(\mathbf{S}^-, \mathbf{S}^+|\theta, \mathbf{T}), \end{aligned} \quad (16)$$

where  $\mathcal{P}(dG|\mathbf{V}, \mathbf{U}, \mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  is equivalent in distribution to  $\mathcal{P}(dG|\mathbf{X}^*, \mathbf{p}, \theta, \mathbf{T})$  given in (32) and  $\pi(\mathbf{S}^-, \mathbf{S}^+|\theta, \mathbf{T})$  is defined in Theorem 2.1(i).

The above characterization of the posterior distribution of  $G$  given  $\theta$  for models in (1) that is in terms of two  $\mathbf{S}$ -paths is less complex than (or as complex as only when  $n, N - n < 4$ ) the partition-based characterization (32) (see Remark 2.7 for discussion in detail). A proof of the above two theorems is given in the Appendix.

Given any path  $\mathbf{S}^+$  and  $\mathbf{S}^-$  of  $n + 1$  and  $N - n + 1$  coordinates, respectively, define

$$\eta_0(\mathbf{S}^+, \mathbf{S}^-) = \frac{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-), 1)}{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))}, \quad (17)$$

for  $j = 1, \dots, n$ ,

$$\eta_j^+(\mathbf{S}^+, \mathbf{S}^-) = \frac{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+) \setminus \{S_j^+ - S_{j-1}^+\}, \mathcal{M}_{1,N-n}(\mathbf{S}^-), S_j^+ - S_{j-1}^+ + 1)}{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))}, \quad (18)$$

if  $S_j^+ > S_{j-1}^+$ , 0 otherwise, and, for  $j = 1, \dots, N - n$ ,

$$\eta_j^-(\mathbf{S}^+, \mathbf{S}^-) = \frac{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-) \setminus \{S_j^- - S_{j-1}^-\}, S_j^- - S_{j-1}^- + 1)}{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))}, \quad (19)$$

if  $S_j^- > S_{j-1}^-$ , 0 otherwise.

**Corollary 2.3.** Theorems 2.1 and 2.2 imply that a Bayes estimate of the unimodal density (1) is given by the posterior mean given  $\theta$  and  $\mathbf{T}$ ,

$$\mathbb{E}[f(t|G, \theta)|\theta, \mathbf{T}] = \sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T}) \pi(\mathbf{S}^-, \mathbf{S}^+|\theta, \mathbf{T}) \quad (20)$$

where

$$a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T}) = \left[ \eta_0(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,0}^+(t) + \sum_{j=1}^n \eta_j^+(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,j}^+(t|\mathbf{S}^+) \right] \mathbb{I}(t > \theta) \\ + \left[ \eta_0(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,0}^-(t) + \sum_{j=1}^{N-n} \eta_j^-(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,j}^-(t|\mathbf{S}^-) \right] \mathbb{I}(t < \theta), \quad (21)$$

with  $d_{\theta,0}^+(t) = \int_{t-\theta}^{\infty} X^{-1} H(dX)$ ,  $d_{\theta,0}^-(t) = \int_{-\infty}^{t-\theta} (-X)^{-1} H(dX)$ , for  $j = 1, \dots, n$ ,

$$d_{\theta,j}^+(t|\mathbf{S}^+) = \begin{cases} \frac{\int_{\max(t-\theta, Y_j)}^{\infty} U^{-(S_j^+ - S_{j-1}^+ + 1)} H(dU)}{\int_{Y_j}^{\infty} U^{-(S_j^+ - S_{j-1}^+)} H(dU)}, & S_j^+ > S_{j-1}^+, \\ 0, & \text{otherwise,} \end{cases}$$

and, for  $j = 1, \dots, N - n$ ,

$$d_{\theta,j}^-(t|\mathbf{S}^-) = \begin{cases} \frac{\int_{-\infty}^{\min(t-\theta, Z_j)} (-V)^{-(S_j^- - S_{j-1}^- + 1)} H(dV)}{\int_{-\infty}^{Z_j} (-V)^{-(S_j^- - S_{j-1}^-)} H(dV)}, & S_j^- > S_{j-1}^-, \\ 0, & \text{otherwise.} \end{cases}$$

The above Bayes estimate is a weighted average of the function  $a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  with respect to  $\pi(\mathbf{S}^-, \mathbf{S}^+|\theta, \mathbf{T})$ . When  $H$  is defined by (31) and the prior components,  $\eta_0(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,0}^+(t)$  and  $\eta_0(\mathbf{S}^+, \mathbf{S}^-) d_{\theta,0}^-(t)$ , vanish, the function  $a_f$  becomes constant between different ordered observations, which is of the same form as Robertson (1967)'s maximum likelihood estimate of the unimodal density when the mode is known as  $\theta$ .

Dividing the right hand side of (36) by the joint distribution of  $(\mathbf{V}, \mathbf{U}, \mathbf{S}^-, \mathbf{S}^+)$  given  $\theta$  and  $\mathbf{T}$ , given by the last line in (16), yields the following analogue of Lemma 2.1 in Ho (2006a) or Corollary 2.4 in Ho (2006b) which states that given  $(\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$ ,  $\mathbf{p}$  is uniformly distributed over all partitions that can be split into  $\mathbf{p}^+$  and  $\mathbf{p}^-$  corresponding to the given paths  $\mathbf{S}^+$  and  $\mathbf{S}^-$ , respectively. The above estimator for a unimodal density follows from the same argument as in Ho (2006b) to be always less variable than its counterpart in terms of partitions due to (32) as a result of Rao–Blackwellization.

**Corollary 2.4.** Consider models in (1). Suppose  $\mathbf{S}^+|\theta, \mathbf{T} \sim \pi^+(\mathbf{S}^+|\theta, \mathbf{T})$  and  $\mathbf{S}^-|\mathbf{S}^+, \theta, \mathbf{T} \sim \pi^-(\mathbf{S}^-|\mathbf{S}^+, \theta, \mathbf{T})$ . Then, there exists a conditional distribution

$$\pi(\mathbf{p}|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T}) = \frac{1}{|\mathbb{C}_{\mathbf{S}^+}| |\mathbb{C}_{\mathbf{S}^-}|}, \quad \mathbf{p} = \mathbf{p}^+ \cup \mathbf{p}^-, \mathbf{p}^+ \in \mathbb{C}_{\mathbf{S}^+}, \mathbf{p}^- \in \mathbb{C}_{\mathbf{S}^-},$$

where  $|\mathbb{C}_{\mathbf{S}}|$  is given by (10).

Suppose  $\theta$  is not known. Theorems 2.1 and 2.2 yield the conditional density of  $\mathbf{T}$  given  $\theta$  to be proportional to  $\sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} \phi_{\theta}^+(\mathbf{S}^+, \mathbf{T}) \phi_{\theta}^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T})$ . A standard prior-posterior updating operation in which the prior distribution of  $\theta$  is  $\pi(d\theta)$  results in the following theorem.

**Theorem 2.5.** Assume any prior  $\pi(d\theta)$  for  $\theta$ . Then, the posterior distribution of  $\theta$  given  $N$  i.i.d. observations  $\mathbf{T}$  from (1) is characterized by

$$\Pr(\theta \in B|\mathbf{T}) = \int_B \sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} \pi(\mathbf{S}^-, \mathbf{S}^+, d\theta|\mathbf{T}), \quad (22)$$

for any Borel set  $B \in \mathcal{R}$ , where

$$\pi(\mathbf{S}^-, \mathbf{S}^+, d\theta|\mathbf{T}) = \frac{\phi_{\theta}^+(\mathbf{S}^+, \mathbf{T}) \phi_{\theta}^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T}) \pi(d\theta)}{\int_{\mathcal{R}} \sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} \phi_{\theta}^+(\mathbf{S}^+, \mathbf{T}) \phi_{\theta}^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T}) \pi(d\theta)}, \quad (23)$$

with  $\phi_{\theta}^+(\mathbf{S}^+, \mathbf{T})$  and  $\phi_{\theta}^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T})$  defined in (12) and (13), respectively.

Finally, the posterior distribution of the pair  $(G, \theta)$  in (1) can be determined from (5) based on Theorems 2.2 and 2.5 and, hence, the posterior expectation of any functional  $h$  of  $(G, \theta)$  can be expressible in terms of a finite sum over two dependent  $\mathbf{S}$ -paths. In particular, a Bayes estimate of the unknown unimodal density (1) follows by letting  $h(G, \theta) = f(t|G, \theta)$  in (5) and applying Corollary 2.3.

**Theorem 2.6.** Assume any prior  $\pi(d\theta)$  for  $\theta$ . Then, the posterior mean of an unimodal density (1) given  $N$  i.i.d. observations  $\mathbf{T}$  is given by

$$\mathbb{E}[f(t|G, \theta)|\mathbf{T}] = \int_{\mathcal{R}} \sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-} a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T}) \pi(\mathbf{S}^-, \mathbf{S}^+, d\theta|\mathbf{T}), \quad (24)$$

where  $a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  and  $\pi(\mathbf{S}^-, \mathbf{S}^+, d\theta|\mathbf{T})$  are defined in Corollary 2.3 and (23), respectively.

**Remark 2.7.** As the total number of partitions of  $N$  integers, which is the Bell's exponential number  $B_N$ , is roughly equal to  $N!$ , the complexity of the partition-based characterization (32), which relies on the total number of partitions  $\mathbf{p}$  (that is, the number of summands in  $\sum_{\mathbf{p}}$ ), is roughly equal to  $n! \times (N-n)!$ . This quantity is identical to that of Brunner's model, but it has not been pointed out in Brunner (1992). Meanwhile, the complexity of the path-based characterization (16), which is based on the number of summands in the double sum  $\sum_{\mathbf{S}^+} \sum_{\mathbf{S}^-}$ , depends on  $\Lambda_n \times \Lambda_{N-n}$  where  $\Lambda_n$  denotes the total number of  $\mathbf{S}$ -paths of  $n+1$  coordinates. Hence, its complexity is less than that of (32) except when both  $n$  and  $N-n$  are less than 4 because  $\Lambda_n \leq B_n$  for all integers  $n$ , with equality only when  $n < 4$  (Brunner and Lo (1989) and Ho (2002)). Table 1 reveals a ratio between the complexities of (16) and (32) to be as large as 0.02097 when  $N = 20$ . This upper bound on the ratio drops quickly when  $N$  increases; for example, the bound is given by  $0.00013^2 = 1.69 \times 10^{-8}$  when  $N = 40$ .

Table 1: Complexities between path-based and partition-based characterizations, (16) and (32), versus sample sizes  $n$  and  $20 - n$ .

$n$	$20 - n$	$\Lambda_n \times \Lambda_{20-n}$	$B_n \times B_{20-n}$	Ratio in %
10	10	282,105,616	13,450,200,625	2.097
8	12	297,457,160	17,444,291,580	1.705
6	14	353,026,080	38,752,562,366	0.911
4	16	495,007,380	157,202,132,205	0.315
2	18	1,432,916,100	2,046,230,418,477	0.070
0	20	6,564,120,420	51,724,158,235,372	0.013

## 2.1 An illustration with the two-parameter Poisson-Dirichlet process

This section illustrates results obtained so far by selecting an important example of the class of species sampling models (2), namely, the two-parameter Poisson-Dirichlet process (Pitman and Yor (1997)). Write the random measure as  $\mathcal{PD}(H; a, b)$

to indicate that its shape probability is  $H$  and there are two shape parameters  $0 \leq a < 1$  and  $b > -a$ . A Dirichlet process with shape measure  $\theta H$ ,  $\theta > 0$ , corresponds to  $\mathcal{PD}(H; 0, \theta)$ . Selections of  $a = \alpha$  and  $b = 0$  give a normalized stable law with index  $0 < \alpha < 1$ , of which a simple exponential change of measure gives a normalized inverse-Gaussian process considered by Lijoi, Mena and Prünster (2005). Posterior analysis of models in (1) wherein  $G$  is  $\mathcal{PD}(H; a, b)$  follow from the previous discussion with explicit simplifications including  $\ell_{0,k} = (b + N_k a)/(b + k)$  and  $\ell_{j,k} = (e_j - a)/(b + k)$  in (3),

$$\chi(\mathcal{M}_{1,n}(\mathbf{S})) = \frac{\prod_{i=1}^{N(\mathbf{S})} [b + (i-1)a] \prod_{\{j^*|\mathbf{S}\}} \prod_{i=1}^{S_j - S_{j-1} - 1} (i - a)}{\prod_{k=1}^n (b + k - 1)}$$

in (11–12),

$$\frac{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))}{\chi(\mathcal{M}_{1,n}(\mathbf{S}^+))} = \frac{\prod_{i=1}^{N(\mathbf{S}^-)} [b + (N(\mathbf{S}^+) + i - 1)a] \prod_{\{j^*|\mathbf{S}^-\}} \prod_{i=1}^{S_j^- - S_{j-1}^- - 1} (i - a)}{\prod_{k=n+1}^N (b + k - 1)}$$

in (13),  $\eta_0(\mathbf{S}^+, \mathbf{S}^-) = [b + (N(\mathbf{S}^+) + N(\mathbf{S}^-))a]/(b + N)$ ,  $\eta_j^+(\mathbf{S}^+, \mathbf{S}^-) = (S_j^+ - S_{j-1}^+ - a)/(b + N)$ , and  $\eta_j^-(\mathbf{S}^+, \mathbf{S}^-) = (S_j^- - S_{j-1}^- - a)/(b + N)$  in (17–19), respectively.

Last but not least, in Theorem 2.2,  $\mathcal{P}(dG|\mathbf{V}, \mathbf{U}, \mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  is equivalent to

$$\sum_{\{j^*|\mathbf{S}^+\}} \frac{G_j^+}{G^*} \delta_{U_j}(\cdot) + \sum_{\{j^*|\mathbf{S}^-\}} \frac{G_j^-}{G^*} \delta_{V_j}(\cdot) + \left( 1 - \sum_{\{j^*|\mathbf{S}^+\}} \frac{G_j^+}{G^*} - \sum_{\{j^*|\mathbf{S}^-\}} \frac{G_j^-}{G^*} \right) \mathcal{P}^*(\cdot),$$

where  $G_j^+ \stackrel{\text{ind}}{\sim} \text{Gamma}(S_j^+ - S_{j-1}^+ - a)$ ,  $G_j^- \stackrel{\text{ind}}{\sim} \text{Gamma}(S_j^- - S_{j-1}^- - a)$ ,  $G^* = \sum_{\{j^*|\mathbf{S}^+\}} G_j^+ + \sum_{\{j^*|\mathbf{S}^-\}} G_j^- + G$  with  $G \sim \text{Gamma}(b + (N(\mathbf{S}^+) + N(\mathbf{S}^-))a)$ , and all variables are mutually independent of  $\mathcal{P}^* = \mathcal{PD}(H; a, b + (N(\mathbf{S}^+) + N(\mathbf{S}^-))a)$ . Note that the above new expression of (13) is a symmetric function, yet different from that in (12), depending on  $\{S_j^- - S_{j-1}^- : S_j^- > S_{j-1}^-, j = 1, \dots, N - n\}$  only. This allows a straightforward application of the SIP sampler (Algorithm 3.1) in drawing  $\mathbf{S}^-$  given  $\mathbf{S}^+$  when constructing an SIS method (Algorithm 3.4) for models in (1) in the next section.

### 3 Sequential Importance Sampling Schemes

This section introduces a SIS method (Kong, Liu and Wong (1994), Liu and Chen (1998) and Liu, Chen and Wong (1998)) for sampling the triplets  $(\mathbf{S}^-, \mathbf{S}^+, \theta)$  in evaluating/approximating posterior quantities for models in (1), like (22) and (24), which are expressible in terms of finite sums of two dependent  $\mathbf{S}$ -paths. The SIS method is based on yet another novel SIS algorithm, called *sequential importance path (SIP) sampler*, for sampling one single path at a time. The SIP sampler is designed in accordance with choosing trial distributions that mimic the probability kernels for Markov transitions in the accelerated path (AP) sampler introduced in Ho (2002, 2006a, 2006b) that serves the same purpose.

Generally speaking, the SIP sampler or any other existing SIS method allows us to draw an  $\mathbf{S}$ -path of  $n + 1$  coordinates according to a probability distribution

$$\pi(\mathbf{S}) \propto \phi(\mathbf{S}) = |\mathbb{C}_{\mathbf{S}}| \chi(\mathcal{M}_{1,n}(\mathbf{S})) \prod_{\{j^*|\mathbf{S}\}} m^{(S_j - S_{j-1})}(Q_j), \quad (25)$$

where  $|\mathbb{C}_{\mathbf{S}}|$  is defined in (10),  $\chi(\cdot)$  is a symmetric function depending on only its arguments, similar to (11),  $m^{(S_j - S_{j-1})}(Q_j)$  is a finite real-valued function depending on  $S_j - S_{j-1}$  and  $Q_j$  only, and  $Q_1, \dots, Q_n$  is a decreasing/increasing sequence in  $\mathcal{R}$ . An inefficient SIS method proposed by Ho (2002, Section 4) consists of  $n - 1$  recursive determinations of one coordinate of the path  $\mathbf{S}$  at a time in an ascending order conditioning on all previously determined coordinates according to a trial distribution

$$\Pr(S_r = s_r | S_1 = s_1, \dots, S_{r-1} = s_{r-1}) := t_r(s_r | s_1, \dots, s_{r-1}) \propto \phi(\mathbf{s}_{r,s_r}^*) \quad (26)$$

for  $r = 1, \dots, n - 1$ , where  $\mathbf{s}_{r,s_r}^* = (0, s_1, \dots, s_{r-1}, s_r, r + 1)$  is a path of  $r + 2$  coordinates. After step  $n - 1$ , a path  $\mathbf{s} = (0, s_1, s_2, \dots, s_{n-1}, n)$  drawn with probability  $t_{n-1}(\mathbf{s}) = \prod_{r=1}^{n-1} t_r(s_r | s_1, \dots, s_{r-1})$  can then be treated as a Monte Carlo sample from (25) after being properly weighted by an importance sampling weight  $\phi(\mathbf{s})/t_{n-1}(\mathbf{s})$ . However, it turns out that the above scheme is practically not efficient

in evaluating sums over  $\mathbf{S}$ -paths. In general, this is directly related to the discrepancy between the trial distribution  $t_i(\cdot|\cdot)$  in (26) and the true conditional distribution of  $S_r$  given  $S_1, \dots, S_{r-1}$  derivable from the target distribution  $\pi(\mathbf{S})$  (Liu and Chen (1998)). Noticing that the “transition” in (26) is equivalent to determinations of the two increments,  $S_r - S_{r-1}$  and  $r + 1 - S_r$ , of the path at locations  $r$  and  $r + 1$ , respectively, our idea is to replace location  $r + 1$  by some other latter location  $q$ , which parallels the idea of constructing the AP sampler adopted in Ho (2002, 2006a,b) when improving on an inefficient Gibbs chain. Let  $I_0 = 0$  and  $I_n = n$  and denote  $\{I_1, \dots, I_{n-1}\}$  as a random permutation of the integers  $\{1, 2, \dots, n-1\}$ , such that  $D_r = \{I_0\} \cup \{I_1, \dots, I_r\} \cup \{I_n\}$  consists of all determined coordinates of the  $\mathbf{S}$ -path after step  $r$  of the SIP sampler, for  $r = 1, \dots, n-1$ ,

**Algorithm 3.1 (Sequential importance path (SIP) sampler).** An efficient SIS method for sampling an  $\mathbf{S}$ -path of  $n + 1$  coordinates from  $\pi(\mathbf{S})$  given in (25), the SIP sampler, consists of recursive applications of the following SIS steps for  $r = 1, \dots, n-1$ :

- A. Given  $D_{r-1}$ , let  $p = \max\{I_j \in D_{r-1} : I_j < I_r\}$  and  $q = \min\{I_j \in D_{r-1} : I_j > I_r\}$ . Determine  $S_{I_r} = k$ , for  $k = S_p, S_p + 1, \dots, \min(I_r, S_q)$ , according to a distribution

$$\kappa_r(k|\{S_h : h \in D_{r-1}\}) \propto \phi(\mathbf{S}_{I_r,k}^*), \quad (27)$$

where  $\mathbf{S}_{I_r,k}^* = (0, S_1^*, \dots, S_{I_r-1}^*, S_{I_r}^*, S_{I_r+1}^*, \dots, S_{n-1}^*, n)$  is a path of  $n + 1$  coordinates such that  $S_{I_r}^* = k$  and for  $i = 1, \dots, I_r - 1, I_r + 1, \dots, n-1$ ,  $S_i^* = S_{I_h}$  if  $i = I_h \in D_{r-1}$ ; otherwise,  $S_i^* = S_{i-1}^*$  (see Remark 3.3 for explicit expressions of  $\kappa_r(k|\{S_h : h \in D_{r-1}\})$  for different values of  $k$ ).

- B. Compute  $\kappa_r(k|\{S_h : h \in D_{r-1}\})$ , equals  $\phi(\mathbf{S}_{I_r,k}^*)$  multiplied by the appropriate constant of proportionality, for the chosen value  $k$  of  $S_{I_r}$ .

After step  $n-1$ , we obtain a random path  $\mathbf{S} = (0, S_1, S_2, \dots, S_{n-1}, n)$  distributed as the trial distribution

$$\kappa_{n-1}(\mathbf{S}) = \prod_{r=1}^{n-1} \kappa_r(S_{I_r} | \{S_h : h \in D_{r-1}\}). \quad (28)$$

Hence, its importance sampling weight is given by  $w_{n-1}(\mathbf{S}) = \phi(\mathbf{S})/\kappa_{n-1}(\mathbf{S})$ . Given  $M$  i.i.d. draws,  $\mathbf{S}_{(1)}, \dots, \mathbf{S}_{(M)}$  with respective importance sampling weights  $w_{n-1}(\mathbf{S}_{(1)}), \dots, w_{n-1}(\mathbf{S}_{(M)})$ , from the SIP sampler based on different permutations  $\{I_1, \dots, I_{n-1}\}$  of the  $n-1$  integers, any sum over  $\mathbf{S}$ -paths/expectation of any functional  $h(\mathbf{S})$  with respect to the probability distribution  $\pi(\mathbf{S})$ ,  $\eta_h = \sum_{\mathbf{S}} h(\mathbf{S})\pi(\mathbf{S})$ , can be approximated by

$$\eta_h^M = \frac{\sum_{i=1}^M h(\mathbf{S}_{(i)}) w_{n-1}(\mathbf{S}_{(i)})}{\sum_{i=1}^M w_{n-1}(\mathbf{S}_{(i)})}. \quad (29)$$

**Remark 3.2.** We remark that there are two major differences between the SIP sampler and the inefficient SIS method which intuitively explain why the SIP sampler is more efficient. On one hand, the coordinates of the path are determined in a random order in the SIP sampler, but not in an ascending order or any other pre-determined order. This arrangement is desired and crucial, as it results in determination of an increment at a location possibly latter than  $r+1$  in step  $r$ , which is the idea behind the success of the AP sampler. On the other hand, each trial distribution  $\kappa_r(S_{I_r} | \{S_h : h \in D_{r-1}\})$  in the SIP sampler mimics the transition probabilities in the AP sampler, in the sense that it is proportional to the probability of a path of  $n+1$  coordinates for any  $r = 1, \dots, n-1$ , rather than the probability of a path of number of coordinates varying with  $r$ .

**Remark 3.3.** In the SIP sampler (Algorithm 3.1), the trial distribution  $\kappa_r(k | \{S_h : h \in D_{r-1}\})$  is explicitly proportional to

$$\frac{I_r - S_p}{S_q - S_p - 1} \chi(\mathcal{M}_{1,p}(\mathbf{S}_{I_r, S_p}^*), S_q - S_p, \mathcal{M}_{q+1,n}(\mathbf{S}_{I_r, S_p}^*)) m^{(S_q - S_p)}(Q_q)$$

if  $k = S_p$ , or



$$\mathbb{I}(I_r \geq S_q) \frac{I_r - S_p}{S_q - S_p - 1} \prod_{i=S_p+1}^{S_q-1} \left( \frac{I_r - i}{q - i} \right) \\ \times \chi(\mathcal{M}_{1,p}(\mathbf{S}_{I_r, S_q}^*), S_q - S_p, \mathcal{M}_{q+1,n}(\mathbf{S}_{I_r, S_q}^*)) m^{(S_q - S_p)}(Q_p)$$

if  $k = S_q$ , or

$$\left( \frac{S_q - S_p - 2}{S_q - k - 1} \right) \prod_{i=I_r+1}^{q-1} \left( \frac{i - k}{I_r - S_p} \right) \\ \times \chi(\mathcal{M}_{1,p}(\mathbf{S}_{I_r, k}^*), k - S_p, S_q - k, \mathcal{M}_{q+1,n}(\mathbf{S}_{I_r, k}^*)) m^{(k - S_p)}(Q_p) m^{(S_q - k)}(Q_q).$$

if  $k = S_p + 1, \dots, \min(I_r, S_q - 1)$ .

**Algorithm 3.4.** An SIS method that samples  $(\mathbf{S}^-, \mathbf{S}^+, \theta)$  from (23) consists of three major steps:

- (i) Sample  $\theta$  according to a density  $\rho(\theta) > 0$ ,  $\theta \in \mathcal{R}$ . Then, define  $\mathbf{Y}$  and  $\mathbf{Z}$  accordingly based on (6). Also, choose random permutations of  $\{1, \dots, n-1\}$  and  $\{1, \dots, N-n-1\}$ .
- (ii) Given  $\theta$ , determine  $\mathbf{S}^+$  by applying Algorithm 3.1 with function  $\phi(\mathbf{S})$  defined by  $\phi_\theta^+(\mathbf{S}^+, \mathbf{T})$  in (12). Obtain  $\kappa_{n-1}(\mathbf{S}^+|\theta)$  according to (28).
- (iii) Given  $(\mathbf{S}^+, \theta)$ , determine  $\mathbf{S}^-$  by applying Algorithm 3.1 with function  $\phi(\mathbf{S})$  defined by  $\phi_\theta^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T})$  in (13), provided that the ratio  $\chi(\mathcal{M}_{1,n}(\mathbf{S}^+), \mathcal{M}_{1,N-n}(\mathbf{S}^-))/\chi(\mathcal{M}_{1,n}(\mathbf{S}^+))$  is a symmetric function of  $\mathcal{M}_{1,N-n}(\mathbf{S}^-)$ . Obtain  $\kappa_{N-n-1}(\mathbf{S}^-|\mathbf{S}^+, \theta)$  according to (28).

After a total of  $N-1$  sub-steps, we obtain a random sample of  $(\mathbf{S}^-, \mathbf{S}^+, \theta)$  distributed as the trial distribution  $\kappa_{N-n-1}(\mathbf{S}^-|\mathbf{S}^+, \theta) \times \kappa_{n-1}(\mathbf{S}^+|\theta) \times \rho(\theta)$ . If  $\pi(d\theta) = \pi(\theta)d\theta$ , its importance sampling weight is given by

$$w_{N-1}(\mathbf{S}^-, \mathbf{S}^+, \theta) = \frac{\phi_\theta^+(\mathbf{S}^+, \mathbf{T}) \phi_\theta^-(\mathbf{S}^-, \mathbf{S}^+, \mathbf{T}) \pi(\theta)}{\kappa_{N-n-1}(\mathbf{S}^-|\mathbf{S}^+, \theta) \kappa_{n-1}(\mathbf{S}^+|\theta) \rho(\theta)}.$$

We remark that it is possible, indeed more desired in terms of efficiency of the SIS method, that the sequence in sampling the two paths in steps (ii) and (iii) can be randomized based on appropriate, but slight, modifications of the function  $\phi(\mathbf{S})$  in applying Algorithm 3.1. That is, there is one-half probability that  $\mathbf{S}^+$  is sampled before  $\mathbf{S}^-$  as stated in Algorithm 3.4; otherwise,  $\mathbf{S}^-$  is sampled before  $\mathbf{S}^+$ .

**Corollary 3.5.** Posterior quantities for models in (1), like (22) and (24), which are expressible as

$$\gamma_h = \int_{\mathcal{R}} \sum_{\mathbf{S}^-} \sum_{\mathbf{S}^+} h(\mathbf{S}^-, \mathbf{S}^+, \theta) \pi(\mathbf{S}^-, \mathbf{S}^+, d\theta | \mathbf{T})$$

can be approximated by

$$\gamma_h^M = \frac{\sum_{i=1}^M h(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)}) w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})}{\sum_{i=1}^M w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})}, \quad (30)$$

where  $(\mathbf{S}_{(1)}^-, \mathbf{S}_{(1)}^+, \theta_{(1)}), \dots, (\mathbf{S}_{(M)}^-, \mathbf{S}_{(M)}^+, \theta_{(M)})$  is a sequence of  $M$  i.i.d. samples from (23) with respective importance sampling weights  $w_{N-1}(\mathbf{S}_{(1)}^-, \mathbf{S}_{(1)}^+, \theta_{(1)}), \dots, w_{N-1}(\mathbf{S}_{(M)}^-, \mathbf{S}_{(M)}^+, \theta_{(M)})$ , obtained by carrying out Algorithm 3.4 independently for a large number of times  $M$ .

## 4 Numerical Results

This section concerns practical applications of our methodology. For purpose of illustration,  $G$  is selected to be the two-parameter Poisson-Dirichlet process as the corresponding results are discussed in Section 2.1. The idea of conjugacy suggests  $H(\cdot)$  of the measure  $\mathcal{PD}(H; a, b)$  to be related to a Pareto distribution. In particular, we chose the following mixture of two Pareto random variables, symmetrical about zero, that is,

$$H(dX) = \frac{\alpha \delta^\alpha}{2(-X)^{\alpha+1}} \mathbb{I}(X < -\delta) dX + \frac{\alpha \delta^\alpha}{2X^{\alpha+1}} \mathbb{I}(X > \delta) dX, \quad \alpha, \delta > 0, \quad (31)$$

such that it results in

$$\int_Y^\infty X^{-\nu} H(dX) = \int_{-\infty}^{-Y} (-X)^{-\nu} H(dX) = \frac{\alpha \delta^\alpha}{2(\alpha + \nu) \max(|Y|, \delta)^{\alpha+\nu}}.$$

for  $Y > 0$  and any positive integer  $\nu$ , which are necessary in implementation of Algorithm 3.4 (or Algorithm 3.1). For purpose of “deflating” the prior belief, we choose  $\alpha = \delta = 0.000001$ . Due to the same reason, the prior  $\pi(d\theta)$  is chosen to be uniformly distributed on a reasonably large interval on  $\mathcal{R}$  such that all observations are included. The sequence in which the coordinates of the  $\mathbf{S}$ -paths are determined, say,  $\{I_1, \dots, I_{n-1}\}$  for a path of  $n+1$  coordinates, is randomized in every application of the sequential algorithms. Likewise, the determinations of the two paths are also randomized in Algorithm 3.4. Last but not least, the Monte Carlo size  $M = 1000$ .

#### 4.1 Resolution of the SIP sampler

This section addresses the performance of the SIP sampler which directly affects the SIS method (Algorithm 3.4) for estimating a unimodal density. Based on a fixed and known mode  $\theta_0$ , our interest is to estimate the unimodal density (1) with  $G$  taken to be the two-parameter Poisson-Dirichlet process with  $H$  in (31),  $a = 0$  and  $b = 1$ , given as in (20) with  $a_f(t|\mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  defined by  $\theta = \theta_0$  together with the simplifications discussed in Section 2.1. To approximate the posterior mean, steps (ii) and (iii) in Algorithm 3.4<sup>2</sup>, which are essentially two sequential applications of the SIP sampler, are implemented based on the known mode  $\theta_0$ . In particular, the convergence property of the approximated density estimate as the sample size  $N$  increases is studied.

Based on nested samples of sizes  $N = 500, 1000$  and  $3000$  from a unimodal

---

<sup>2</sup>As discussed after the introduction of Algorithm 3.4, the sequence of determinations of the two paths –  $\mathbf{S}^+$  first or  $\mathbf{S}^-$  first – is randomized to achieve a higher efficiency.

density with  $[-1, 0]$  as modal interval (Wegman 1970a) given by

$$\lambda_1(t) = \begin{cases} 0.02 & -7 < t \leq -2 \\ 0.1 & -2 < t \leq -1 \\ 0.4 & -1 < t \leq 0 \\ 0.4 \exp(-t) & t > 0 \\ 0 & \text{otherwise,} \end{cases}$$

weighted averages that approximate the posterior mean of the unimodal density conditioning on  $\theta = \theta_0$ , given as in (30),

$$\gamma_{af}^M(t|\theta_0) = \frac{\sum_{i=1}^M a_f(t|\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_0, \mathbf{T}) w_{N-2}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+|\theta_0)}{\sum_{i=1}^M w_{N-2}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+|\theta_0)},$$

where  $w_{N-2}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+|\theta_0)$  is the importance sampling weight of the pair  $(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+)$  resulted from steps (ii) and (iii) of Algorithm 3.4, are displayed at the left columns in Figures 1-3 for  $\theta_0 = -1, -0.5$  (center mode), and 0, respectively. The whole procedure is repeated for the two-parameter Poisson-Dirichlet process with  $a = 0.9$  and  $b = 100$ . The density estimates based on the three selected values of the mode are depicted at the right columns in Figures 1-3. The graphs echo the fact that the approximated Bayes estimate of the unimodal density,  $\gamma_{af}^M(t|\theta_0)$ , tends to the “true” unimodal density  $\lambda_1(t)$  as sample size increases (from top to bottom in the figures) regardless of the two sets of parameters for  $G$  (between columns in the figures). When  $N$  is large, there is not much difference among density estimates based on different modes.

## 4.2 Resolution of the SIS method (Algorithm 3.4)

The practicality of the SIS method (Algorithm 3.4) for estimation of a unimodal density and its mode is addressed in this section. To estimate the unimodal density (1) with  $G$  taken to the two-parameter Poisson-Dirichlet process with  $a = 0$  and  $b = 1$ , Algorithm 3.4 based on  $\rho(\theta)$  as a standard normal density is implemented independently for  $M = 1000$  number of times to produce random samples of  $(\mathbf{S}^-, \mathbf{S}^+, \theta)$

with importance sampling weight  $w_{N-1}(\mathbf{S}^-, \mathbf{S}^+, \theta)$ . According to Corollary 3.5, two weighted averages, defined as in (30),

$$\theta^M = \frac{\sum_{i=1}^M \theta_{(i)} w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})}{\sum_{i=1}^M w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})}$$

and

$$\gamma_{a_f}^M(t) = \frac{\sum_{i=1}^M a_f(t | \mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)}, \mathbf{T}) w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})}{\sum_{i=1}^M w_{N-1}(\mathbf{S}_{(i)}^-, \mathbf{S}_{(i)}^+, \theta_{(i)})},$$

are used to approximate the Bayes estimates (the posterior mean given  $N$  observations) of the unknown mode  $\theta$  and the unknown unimodal density, respectively.

The unimodal density  $\lambda_1(t)$  in the previous section and another two unimodal densities are chosen as test densities. They are,

$$\lambda_2(t) = \begin{cases} 0.02 & -7 < t \leq -2 \\ 0.25 & -2 < t \leq 0 \\ 0.5 & 0 < t \leq 0.1 \\ 0.1 & 0.1 < t \leq 2.5 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\lambda_3(t) = \frac{12}{13} \left[ \zeta(1.5x) \mathbb{I}(-\infty < x < 0) + \zeta\left(\frac{x}{1.5}\right) \mathbb{I}(0 < x < \infty) \right],$$

where  $\zeta(\cdot)$  is the density function of a standard Cauchy random variable. These three densities behave quite differently from one another in the sense that they have modal interval of length 1, modal interval of shorter length 0.1, and a unique mode at zero, respectively.

Density estimates  $\gamma_{a_f}^M(t)$  based on nested samples of sizes  $N = 500, 1000$  and  $2000$  from the three unimodal densities are depicted in the left columns of Figures 4-6, respectively, while mode estimates  $\theta^M$  are presented in Table 2. The whole procedure is repeated with  $\rho(\theta)$  as a less diffuse normal density with mean 0 and standard deviation  $1/4$ . The resulting density estimates are depicted in the right columns of Figures 4-6, while mode estimates are appended in Table 2. It is evident from

the mode estimates in Table 2, especially when  $N$  is not large, that approximation results based on  $\rho(\theta)$  with a smaller standard deviation are much better than the others. This is also supported by Figures 4-6; for instance, the peak at the modal interval  $[0, 0.1]$  is not well-captured even when  $N = 2000$  (graph at the bottom-left in Figure 5). This phenomenon can be addressed by the well-known fact (Kong, Liu and Wong (1994) and Liu, Chen and Wong (1998)) that efficiency of any SIS method depends heavily on whether the initial trial distributions in its early steps/stages is close to the true conditional distributions. Hence, a good choice of  $\rho(\theta)$  in step (i) of Algorithm 3.4 directly affects the efficiency of the SIS method.

Table 2: Weighted average estimates of the mode

$\rho(\theta)$	$N$	Unimodal Density		
		$\lambda_1(t)$	$\lambda_2(t)$	$\lambda_3(t)$
$N(0, 1)$	500	-1.249450	-1.538695	-0.230042
	1000	1.068161	0.079308	0.815681
	2000	-0.999335	0.037052	-0.615350
$N(0, 0.25^2)$	500	0.165998	0.138150	-0.071292
	1000	0.199645	0.143027	0.013269
	2000	0.101668	-0.071598	-0.271294
True mode		$[-1, 0]$	$[0, 0.1]$	0

To explore the selection issue of  $\rho(\theta)$ , we carry out a large-sample study by replicating the above procedure to estimate the mode of the unimodal density  $\lambda_3(t)$  based on  $N = 500$  observations. Histograms of the 2000 independent Bayes estimates of  $\theta$  based on the two different  $\rho(\theta)$ 's are plotted in Figure 7. It is clear from the graph in the last row based on a standard normal density for  $\rho(\theta)$  does not give convincing posterior estimates of the mode. On the contrary, the graph in the second row shows that the true mode is well-captured when  $\rho(\theta)$  is less diffuse. This deficiency can be understood by looking at the histogram of the 500 observations in the first row

in Figure 7; the posterior distribution of the mode  $\theta$  should be quite concentrated around zero and, hence, the choice of a standard normal density for  $\rho(\theta)$  may be far too diffuse. Note that regarding the results for the first two test densities, the less diffuse normal density with standard deviation  $1/4$ , symmetrical about zero, is not really close to the posterior distribution of  $\theta$  at all based on the histograms of the data in Figure 8. This implies that it is not necessary to set  $\rho(\theta)$  to be extremely close to the true posterior distribution of  $\theta$ , which is characterized in Theorem 2.5. In conclusion, we suggest setting  $\rho(\theta)$  to be a density which is not too diffuse around the mode (based on information from the histogram of the data) in applying the SIS method (Algorithm 3.4) for estimating unimodal densities.

## Appendix: Proof of Theorem 2.1 and 2.2

*Proof.* Suppose  $\theta$  is given. Theorem 2 in Ishwaran and James (2003) states that the law of  $G$  in (1) given  $\theta$  and  $N$  i.i.d. observations  $\mathbf{T}$  is characterized by

$$\begin{aligned} & \int_{\mathcal{M}} g(G) \mathcal{P}(dG|\theta, \mathbf{T}) \\ &= \sum_{\mathbf{p}} \left[ \int_{\mathcal{R}^{N(\mathbf{p})}} \left\{ \int_{\mathcal{M}} g(G) \mathcal{P}(dG|\mathbf{X}^*, \mathbf{p}, \theta, \mathbf{T}) \right\} \prod_{k=1}^{N(\mathbf{p})} \mu(dX_k^*|C_k) \right] \pi(\mathbf{p}|\theta, \mathbf{T}) \end{aligned} \quad (32)$$

for any nonnegative or integrable function  $g$ , wherein  $\mathcal{P}(dG|\mathbf{X}^*, \mathbf{p}, \theta, \mathbf{T})$  is determined by Lemma 1 in Ishwaran and James (2003), and  $\prod_{k=1}^{N(\mathbf{p})} \mu(dX_k^*|C_k) \pi(\mathbf{p}|\theta, \mathbf{T})$  is equivalent in distribution to the posterior distribution of  $\mathbf{X}$  given  $\theta$  as discussed in Theorem 1 in Ishwaran and James (2003), where, for  $k = 1, \dots, N(\mathbf{p})$ ,  $\mu(dX_k^*|C_k)$  is proportional to

$$\varphi_k(dX_k^*) = \frac{1}{(X_k^*)^{e_k}} \left[ \mathbb{I}(0 < \max_{j \in C_k} T_j - \theta \leq X_k^*) - \mathbb{I}(X_k^* \leq \min_{j \in C_k} T_j - \theta < 0) \right] H(dX_k^*) \quad (33)$$

and  $\pi(\mathbf{p}|\theta, \mathbf{T}) = \pi(\mathbf{p}) \prod_{k=1}^{N(\mathbf{p})} \int \varphi_k(dX_k^*) / \sum_{\mathbf{p}} \pi(\mathbf{p}) \prod_{k=1}^{N(\mathbf{p})} \int \varphi_k(dX_k^*)$ .

Splitting  $\mathbf{p}$  into  $\mathbf{p}^+$  and  $\mathbf{p}^-$  as discussed in Section 2 and re-expressing the kernels according to (7) yield

$$\varphi_k(dX_k^*) = \begin{cases} \varphi_k^+(dX_k^*) = (X_k^*)^{-e_k} \mathbb{I}(0 < \max_{j \in C_k} Y_j \leq X_k^*) H(dX_k^*) & k \leq N(\mathbf{p}^+) \\ \varphi_k^-(dX_k^*) = (-X_k^*)^{-e_k} \mathbb{I}(X_k^* \leq \min_{j \in C_k} Z_j < 0) H(dX_k^*) & k > N(\mathbf{p}^+) \end{cases} \quad (34)$$

and

$$\begin{aligned} \pi(\mathbf{p}|\theta, \mathbf{T}) &= \frac{\pi(\mathbf{p}) \left[ \prod_{i=1}^{N(\mathbf{p}^+)} \int \varphi_i^+(dX_i^*) \right] \left[ \prod_{j=N(\mathbf{p}^+)+1}^{N(\mathbf{p})} \int \varphi_j^-(dX_j^*) \right]}{\sum_{\mathbf{p}} \pi(\mathbf{p}) \left[ \prod_{i=1}^{N(\mathbf{p}^+)} \int \varphi_i^+(dX_i^*) \right] \left[ \prod_{j=N(\mathbf{p}^+)+1}^{N(\mathbf{p})} \int \varphi_j^-(dX_j^*) \right]} \\ &= \frac{\psi^-(\mathbf{p}^-|\mathbf{p}^+, \theta, \mathbf{T}) \times \psi^+(\mathbf{p}^+|\theta, \mathbf{T})}{\sum_{\mathbf{p}^+} \left\{ \sum_{\mathbf{p}^-} \psi^-(\mathbf{p}^-|\mathbf{p}^+, \theta, \mathbf{T}) \right\} \psi^+(\mathbf{p}^+|\theta, \mathbf{T})}, \end{aligned} \quad (35)$$

where  $\psi^+(\mathbf{p}^+|\theta, \mathbf{T}) = \pi(\mathbf{p}^+) \prod_{i=1}^{N(\mathbf{p}^+)} \int \varphi_i^+(dX_i^*)$  defines a posterior distribution of  $\mathbf{p}^+$  of the  $n$  positive observations  $\mathbf{Y}$  given  $\theta$  and  $\psi^-(\mathbf{p}^-|\mathbf{p}^+, \theta, \mathbf{T}) = \pi(\mathbf{p}^-|\mathbf{p}^+) \prod_{j=N(\mathbf{p}^+)+1}^{N(\mathbf{p})} \int \varphi_j^-(dX_j^*)$  defines a (conditional) posterior distribution of  $\mathbf{p}^-$  of the remaining  $N - n$  negative observations given  $(\mathbf{p}^+, \theta)$ . The equality in (35) follows from (9) and re-writing  $\sum_{\mathbf{p}} = \sum_{\mathbf{p}^+} \sum_{\mathbf{p}^-}$ . Then, combining (34) and (35) gives

$$\begin{aligned} \prod_{k=1}^{N(\mathbf{p})} \mu(dX_k^*|C_k) \pi(\mathbf{p}|\theta, \mathbf{T}) &\propto \left[ \prod_{j=N(\mathbf{p}^+)+1}^{N(\mathbf{p})} \varphi_j^-(dX_j^*) \psi^-(\mathbf{p}^-|\mathbf{p}^+, \theta, \mathbf{T}) \right] \\ &\quad \times \left[ \prod_{i=1}^{N(\mathbf{p}^+)} \varphi_i^+(dX_i^*) \psi^+(\mathbf{p}^+|\theta, \mathbf{T}) \right]. \end{aligned} \quad (36)$$

Theorem 2.1 in Ho (2006b) yields that the law of  $X_1^*, \dots, X_{N(\mathbf{p}^+)}^*, \mathbf{p}^+|\theta, \mathbf{T}$  (proportional to the last term above) is equivalent to the law of  $\mathbf{U}, \mathbf{S}^+|\theta, \mathbf{T}$  defined by (12) and (14). Utilizing the symmetric properties of  $\pi(\mathbf{p}^-|\mathbf{p}^+)$  in (8) and (11) and applying Theorem 2.1 in Ho (2006b) yield the equivalence between the law of  $X_{N(\mathbf{p}^+)+1}^*, \dots, X_{N(\mathbf{p})}^*, \mathbf{p}^-|\mathbf{p}^+, \theta, \mathbf{T}$ , proportional to the first term at the right hand side of (36), and the law of  $\mathbf{V}, \mathbf{S}^-|\mathbf{S}^+, \theta, \mathbf{T}$  defined by (13) and (15), completing the proof of Theorem 2.1. The result in Theorem 2.2 follows as a result of Theorem 2.1 by recognizing the equality in distribution between  $\mathcal{P}(dG|\mathbf{X}^*, \mathbf{p}, \theta, \mathbf{T})$  and  $\mathcal{P}(dG|\mathbf{V}, \mathbf{U}, \mathbf{S}^-, \mathbf{S}^+, \theta, \mathbf{T})$  in (32).  $\square$



## References

- [1] BERTIN, E. M. J., CUCULESCU, I. and THEODORESCU, R. (1997). *Unimodality of probability measures*. Dordrecht: Kluwer Academic Publishers Group.
- [2] BICKEL, P. J. and FAN, J. (1996). Some problems on the estimation of unimodal densities. *Statist. Sinica* **6** 23–46.
- [3] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981.
- [4] BRUNNER, L. J. (1992). Bayesian nonparametric methods for data from a unimodal density. *Statist. Probab. Lett.* **14** 195–199.
- [5] BRUNNER, L. J. (1995). Bayesian linear regression with error terms that have symmetric unimodal densities. *J. Nonparametr. Statist.* **4** 335–348.
- [6] BRUNNER, L. J. and LO, A. Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.* **17** 1550–1566.
- [7] BRUNNER, L. J. and LO, A. Y. (1994). Nonparametric Bayes methods for directional data. *Canad. J. Statist.* **22** 401–412.
- [8] BUCKLE, D. J. (1995). Bayesian inference for stable distributions. *J. Amer. Statist. Assoc.* **90** 605–613.
- [9] CHENG, M.-Y., GASSER, T. and HALL, P. (1999). Nonparametric density estimation under unimodality and monotonicity constraints. *J. Comput. Graph. Statist.* **8** 1–21.
- [10] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Math. Statist.* **16** 31–41.
- [11] DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications*. New York: Academic Press.

- [12] DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.
- [13] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.
- [14] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications* (Vol. 2, 2nd ed.), New York: John Wiley & Sons.
- [15] FERNÁNDEZ, C. and STEEL, M. F. J. (1998). On Bayesian modelling of fat tails and skewness. *J. Amer. Statist. Assoc.* **93** 359–371.
- [16] FERREIRA, J. T. and STEEL, M. F. J. (2006). A constructive representation of univariate skewed distributions. *J. Amer. Statist. Assoc.* **101** 823–829.
- [17] GRENANDER, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift* **39** 125–153.
- [18] HALL, P. and HUANG, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraint. *Ann. Statist.* **29** 624–647.
- [19] HALL, P. and HUANG, L.-S. (2002). Unimodal density estimation using kernel methods. *Statist. Sinica* **12** 965–990.
- [20] HO, M.-W. (2002). Bayesian inference for models with monotone densities and hazard rates. Unpublished Ph.D. thesis, The Hong Kong University of Science and Technology, Dept. of Information and Systems Management.
- [21] HO, M.-W. (2006a). A Bayes method for a monotone hazard rate via **S**-paths. *Ann. Statist.* **34** 820–836.
- [22] HO, M.-W. (2006b). Bayes estimation of a symmetric unimodal density via **S**-paths. *J. Comput. Graph. Statist.* to appear.

- [23] ISHWARAN, H. and JAMES, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica* **13** 1211–1235.
- [24] ISHWARAN, H. and ZAREPOUR, M. (2002a). Exact and approximate sum-representations for the Dirichlet process. *Canad. J. Statist.* **30** 269–283.
- [25] ISHWARAN, H. and ZAREPOUR, M. (2002b). Dirichlet prior sieves in finite normal mixtures. *Statist. Sinica* **12** 941–963.
- [26] JONES, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test* **13** 1–43.
- [27] KHINTCHINE, A. Y. (1938). On unimodal distributions. *Izvestiya Nauchnolssledovatel'skogo Instituta Matematiki i Mekka* **2** 1–7.
- [28] KONG, A., LIU, J. S. and WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89** 278–288.
- [29] KOTTAS, A. and GELFAND, A. E. (2001). Bayesian semiparametric median regression modeling. *J. Amer. Statist. Assoc.* **96** 1458–1468.
- [30] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291.
- [31] LIU, J. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044.
- [32] LIU, J., CHEN, R. and WONG, W. H. (1998). Rejection control and sequential importance sampling. *J. Amer. Statist. Assoc.* **93** 1022–1031.
- [33] LO, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimation. *Ann. Statist.* **12** 351–357.

- [34] LO, A. Y., BRUNNER, L. J. and CHAN, A. T. (1996). Weighted Chinese restaurant processes and Bayesian mixture models. Research report 1, The Hong Kong University of Science and Technology, Dept. of Information and Systems Management. Available at <http://www.erin.utoronto.ca/~jbrunner/papers/wcr96.pdf>.
- [35] LO, A. Y. and WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245.
- [36] LYE, J. N. and MARTIN, V. L. (1993). Robust estimation, nonnormalities, and generalized exponential distributions. *J. Amer. Statist. Assoc.* **88** 261–267.
- [37] PARZEN, E. (1962). On estimation of a probability density function and its mode. *Ann. Math. Statist.* **33** 1065–1076.
- [38] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158.
- [39] PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (Edited by T. S. Ferguson, L. S. Shapley and J. B. MacQueen). IMS Lecture Notes-Monograph Series (vol. 30), pp. 245–267, Hayward, CA: Institute of Mathematical Statistics.
- [40] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900.
- [41] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā* **31** 23–36.
- [42] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of random measures with independent increments” *Ann. Statist.* **31** 560–585.

- [43] ROBERTSON, T. (1967). On estimating a density which is measurable with respect to a  $\sigma$ -lattice. *Ann. Math. Statist.* **38** 482–493.
- [44] SHEPP, L. A. (1962). Symmetric random walk. *Trans. Amer. Math. Soc.* **104** 144–153.
- [45] VENTER, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38** 1446–1455.
- [46] WANG, Y. (1995). The  $L_1$  theory of estimation of monotone and unimodal densities. *J. Nonparametr. Stat.* **4** 249–261.
- [47] WEGMAN, R. Y. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.* **40** 1661–1667.
- [48] WEGMAN, R. Y. (1970a). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41** 457–471.
- [49] WEGMAN, R. Y. (1970b). Maximum likelihood estimation of a unimodal density II. *Ann. Math. Statist.* **41** 2169–2174.
- [50] WEGMAN, R. Y. (1971). A note on the estimation of the mode. *Ann. Math. Statist.* **42** 1909–1915.

MAN-WAI HO

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

6 SCIENCE DRIVE 2

SINGAPORE 117546

REPUBLIC OF SINGAPORE

E-mail: stahmw@nus.edu.sg

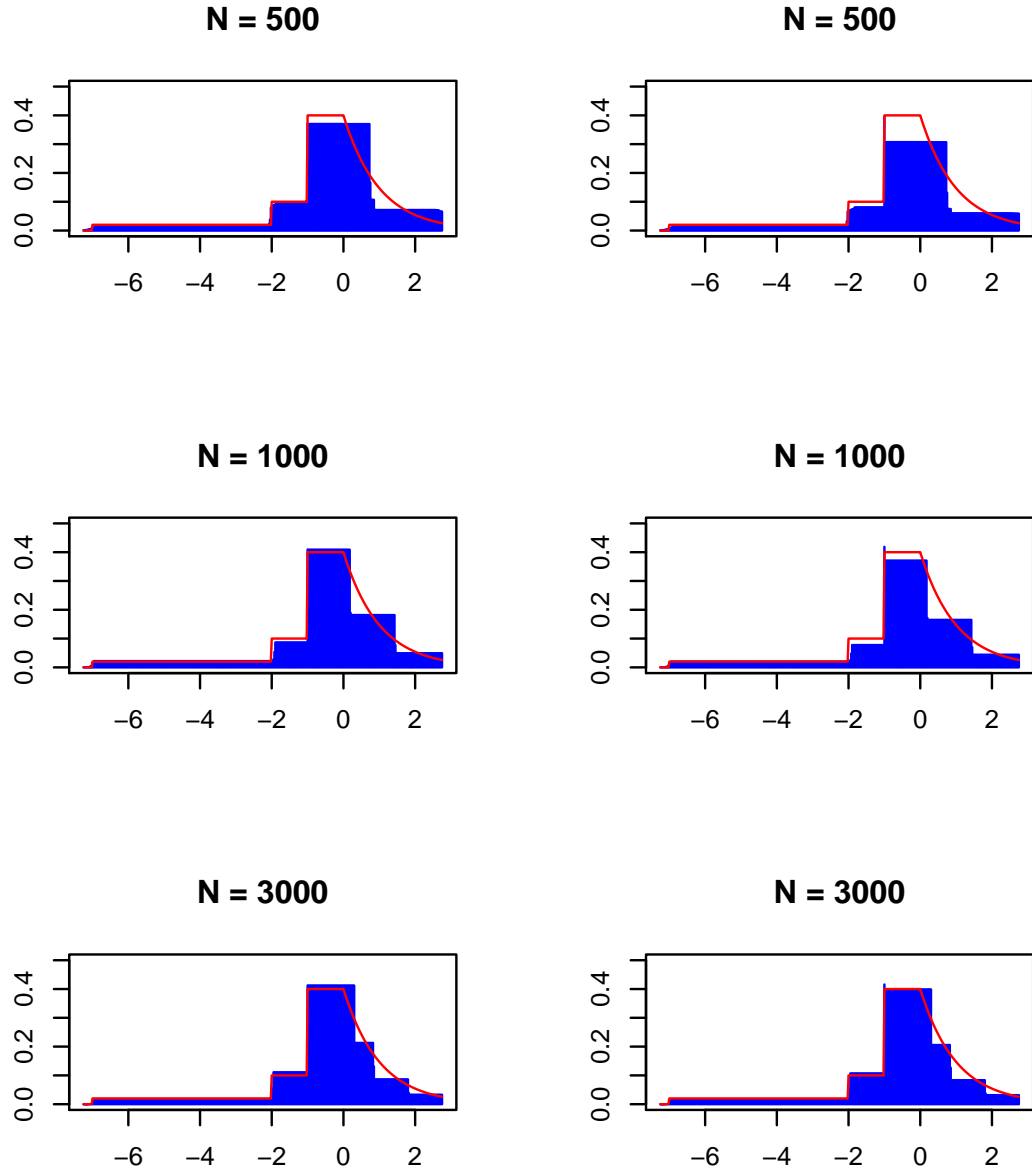


Figure 1: The true unimodal density  $\lambda_1(t)$  (solid line) and weighted average density estimates given the mode  $\theta = -1$  produced by the SIP sampler (Steps (ii) and (iii) of Algorithm 3.4) based on  $a = 0$  and  $b = 1$  (left column) and  $a = 0.9$  and  $b = 100$  (right column) for  $G$ .

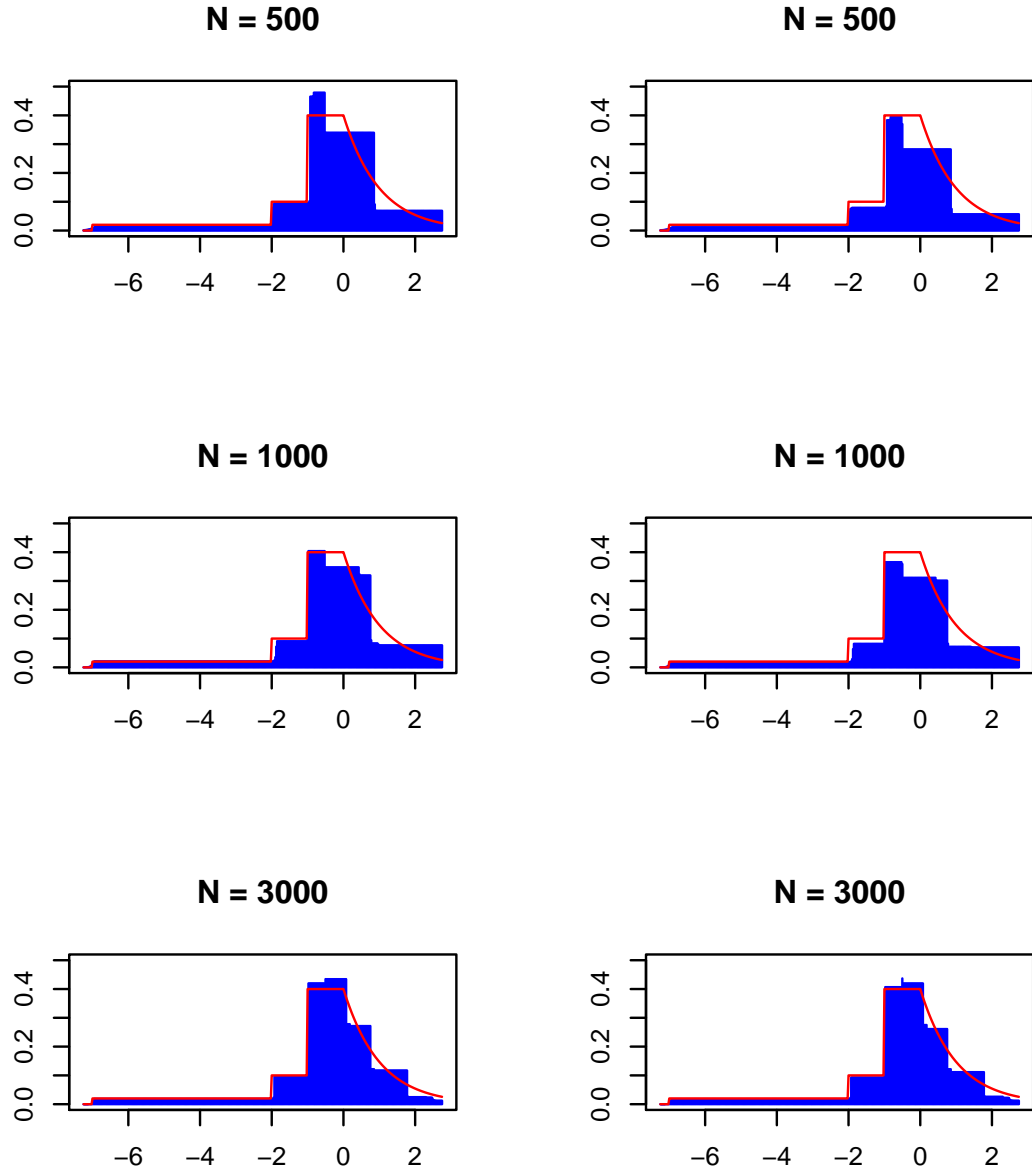


Figure 2: The true unimodal density  $\lambda_1(t)$  (solid line) and weighted average density estimates given the mode  $\theta = -0.5$  produced by the SIP sampler (Steps (ii) and (iii) of Algorithm 3.4) based on  $a = 0$  and  $b = 1$  (left column) and  $a = 0.9$  and  $b = 100$  (right column) for  $G$ .

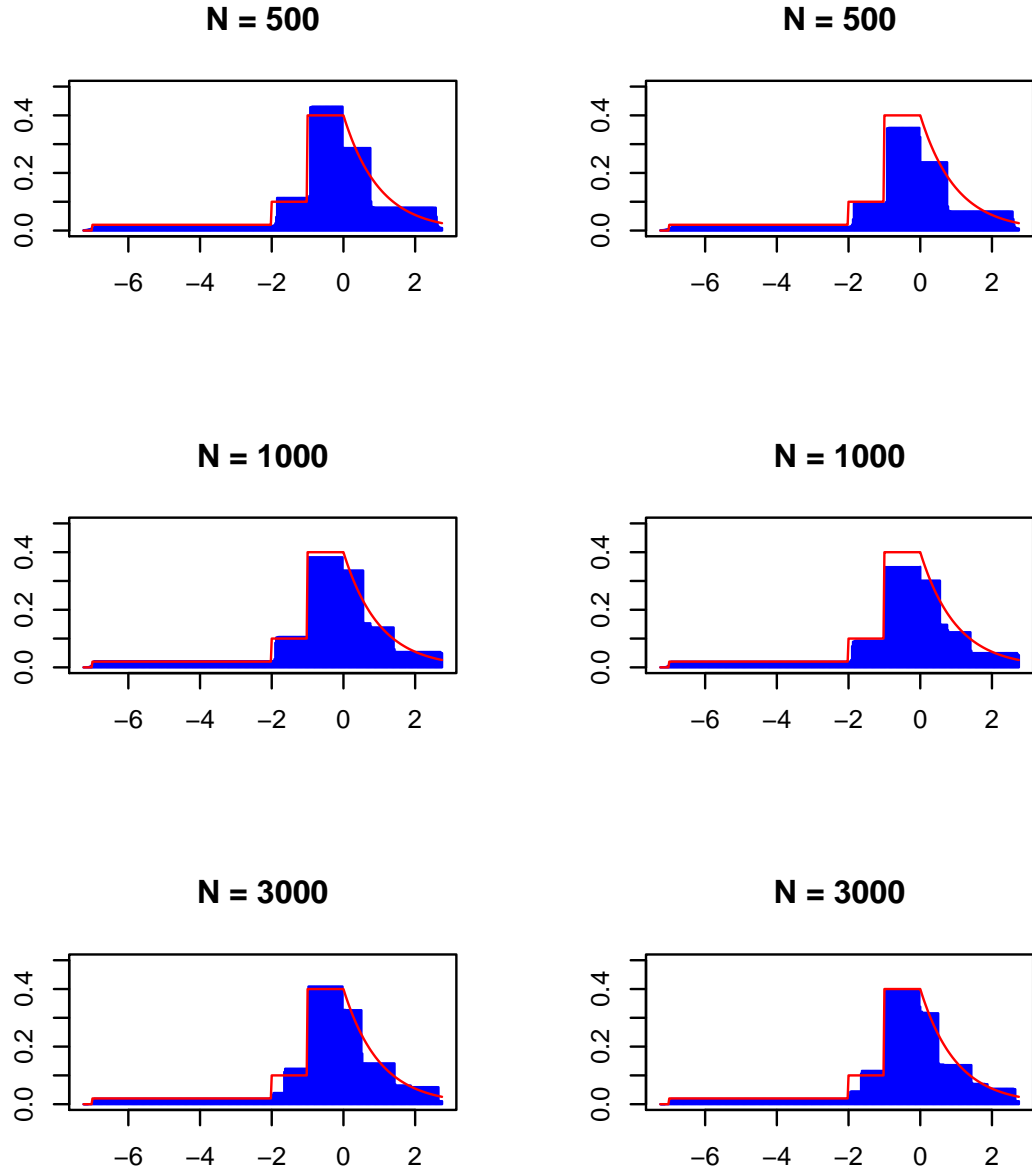


Figure 3: The true unimodal density  $\lambda_1(t)$  (solid line) and weighted average density estimates given the mode  $\theta = 0$  produced by the SIP sampler (Steps (ii) and (iii) of Algorithm 3.4) based on  $a = 0$  and  $b = 1$  (left column) and  $a = 0.9$  and  $b = 100$  (right column) for  $G$ .



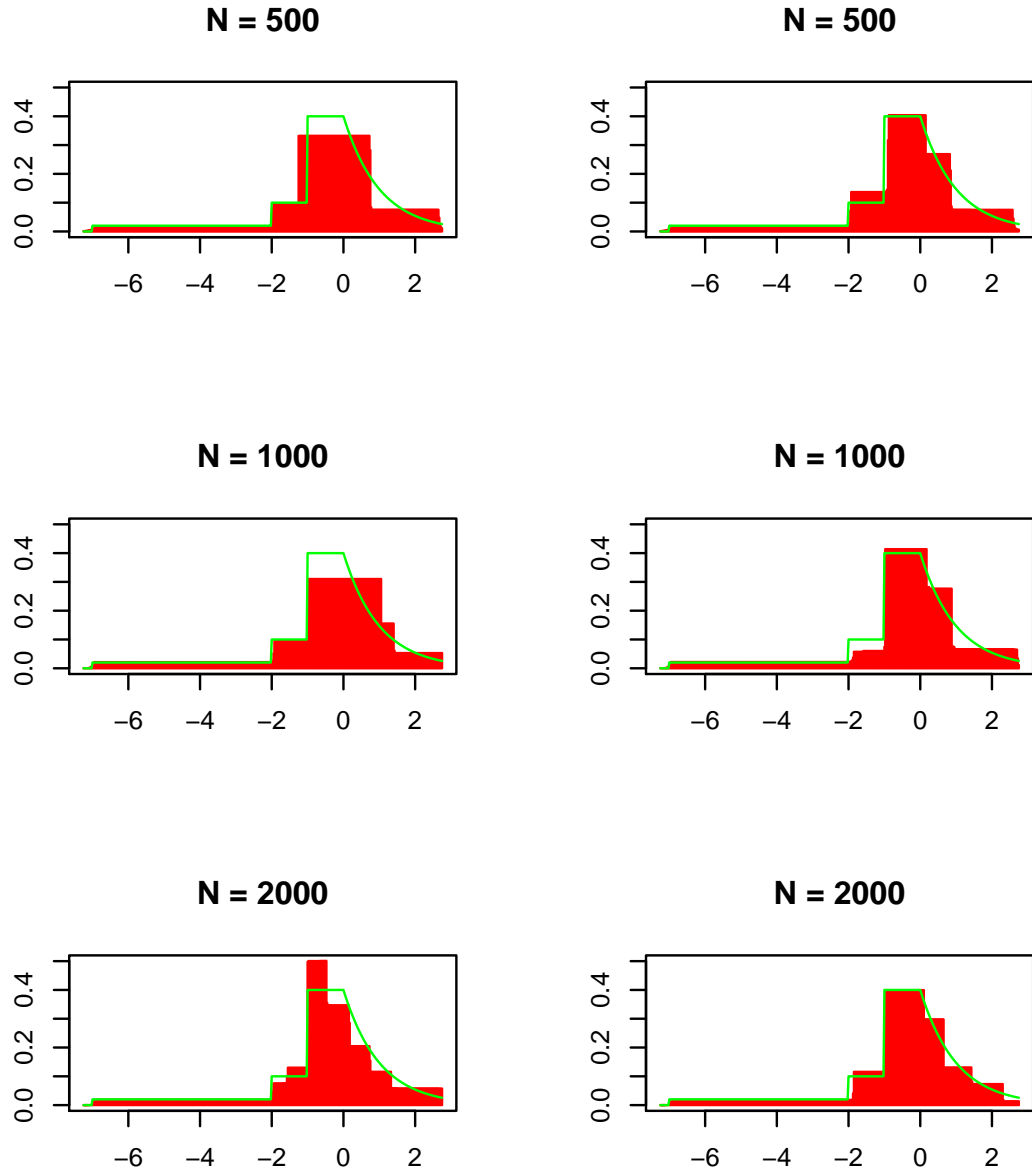


Figure 4: The true unimodal densities  $\lambda_1(t)$  (solid lines) and weighted average density estimates produced by Algorithm 3.4 based on a  $N(0, 1)$  density (left column) and a  $N(0, 0.25^2)$  density (right column) for  $\rho(\theta)$ .

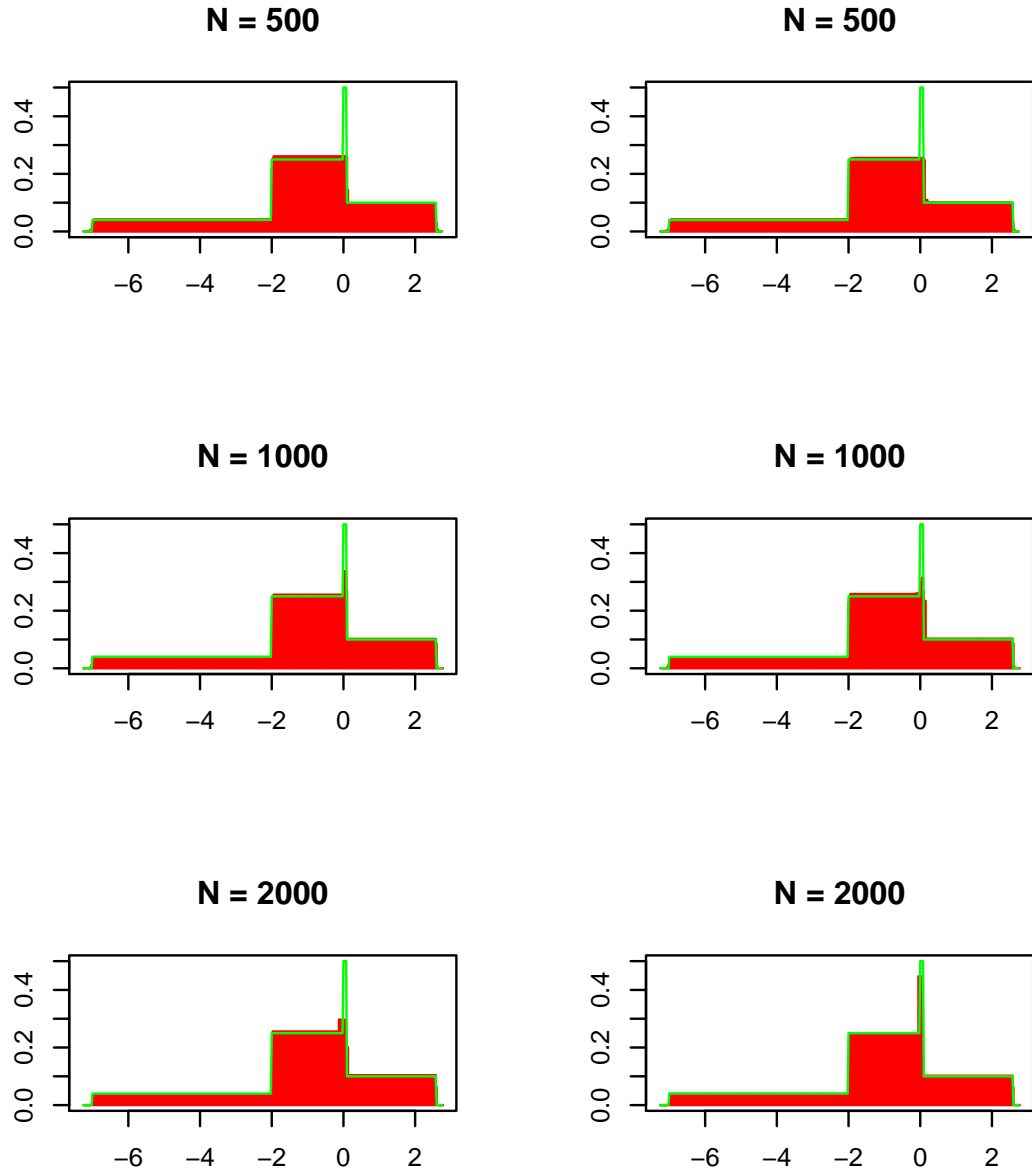


Figure 5: The true unimodal densities  $\lambda_2(t)$  (solid lines) and weighted average density estimates produced by Algorithm 3.4 based on a  $N(0, 1)$  density (left column) and a  $N(0, 0.25^2)$  density (right column) for  $\rho(\theta)$ .

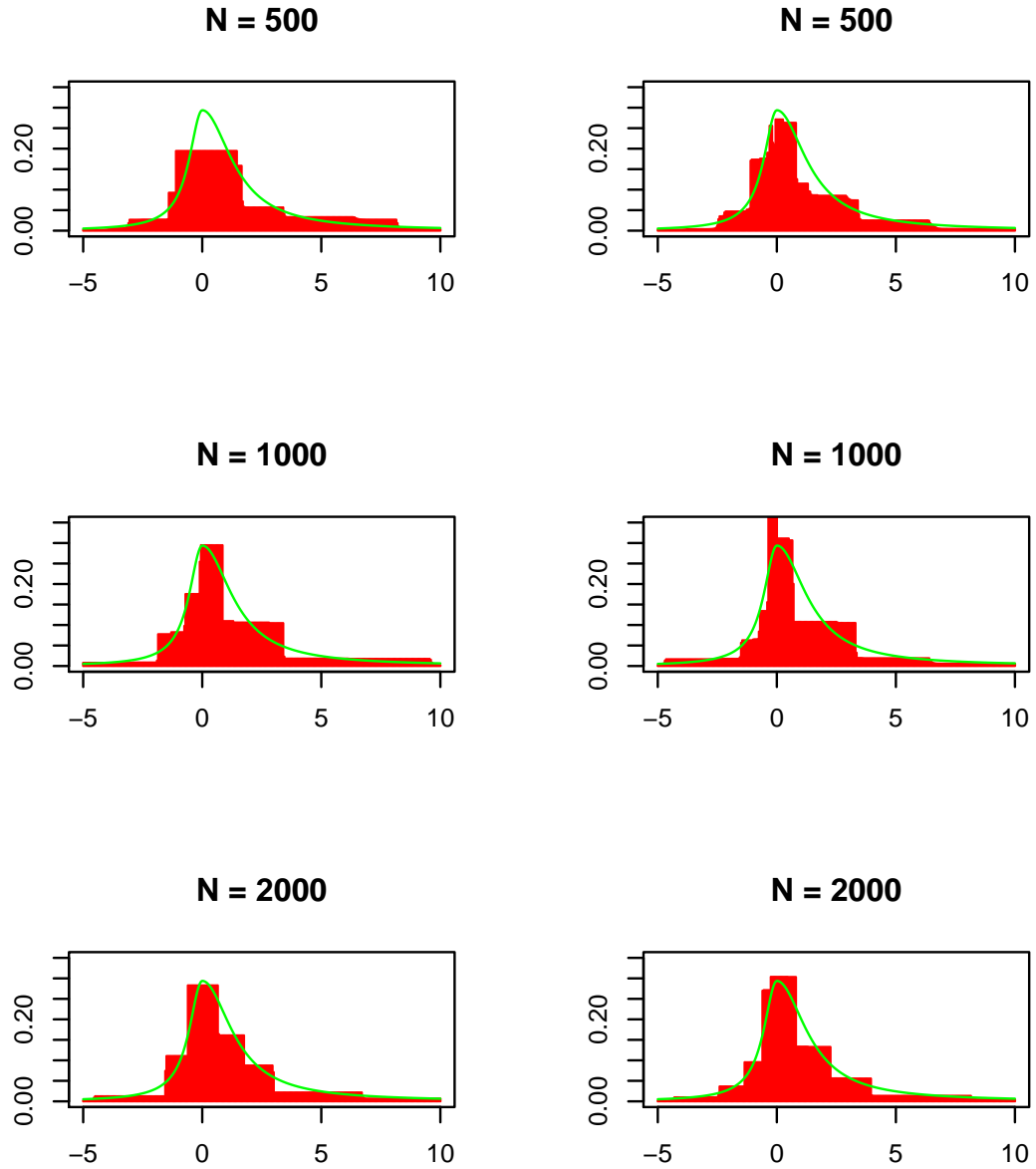


Figure 6: The true unimodal densities  $\lambda_3(t)$  (solid lines) and weighted average density estimates produced by Algorithm 3.4 based on a  $N(0,1)$  density (left column) and a  $N(0, 0.25^2)$  density (right column) for  $\rho(\theta)$ .

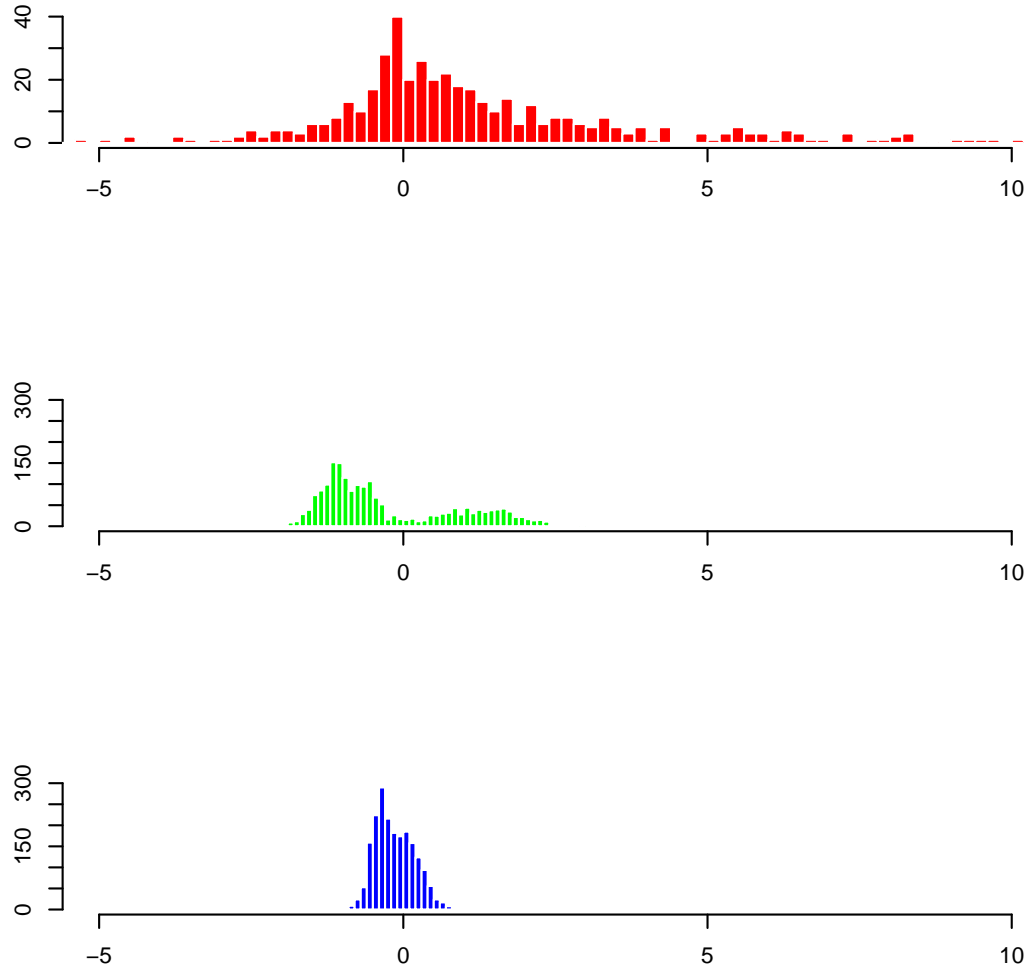


Figure 7: Histogram of 500 observations simulated from  $\lambda_3(t)$  and histograms of the resulting Bayes estimates of  $\theta$  by 2000 replications of Algorithm 3.4 based on a  $N(0, 1)$  density and a  $N(0, 0.25^2)$  density for  $\rho(\theta)$  (from top to bottom).

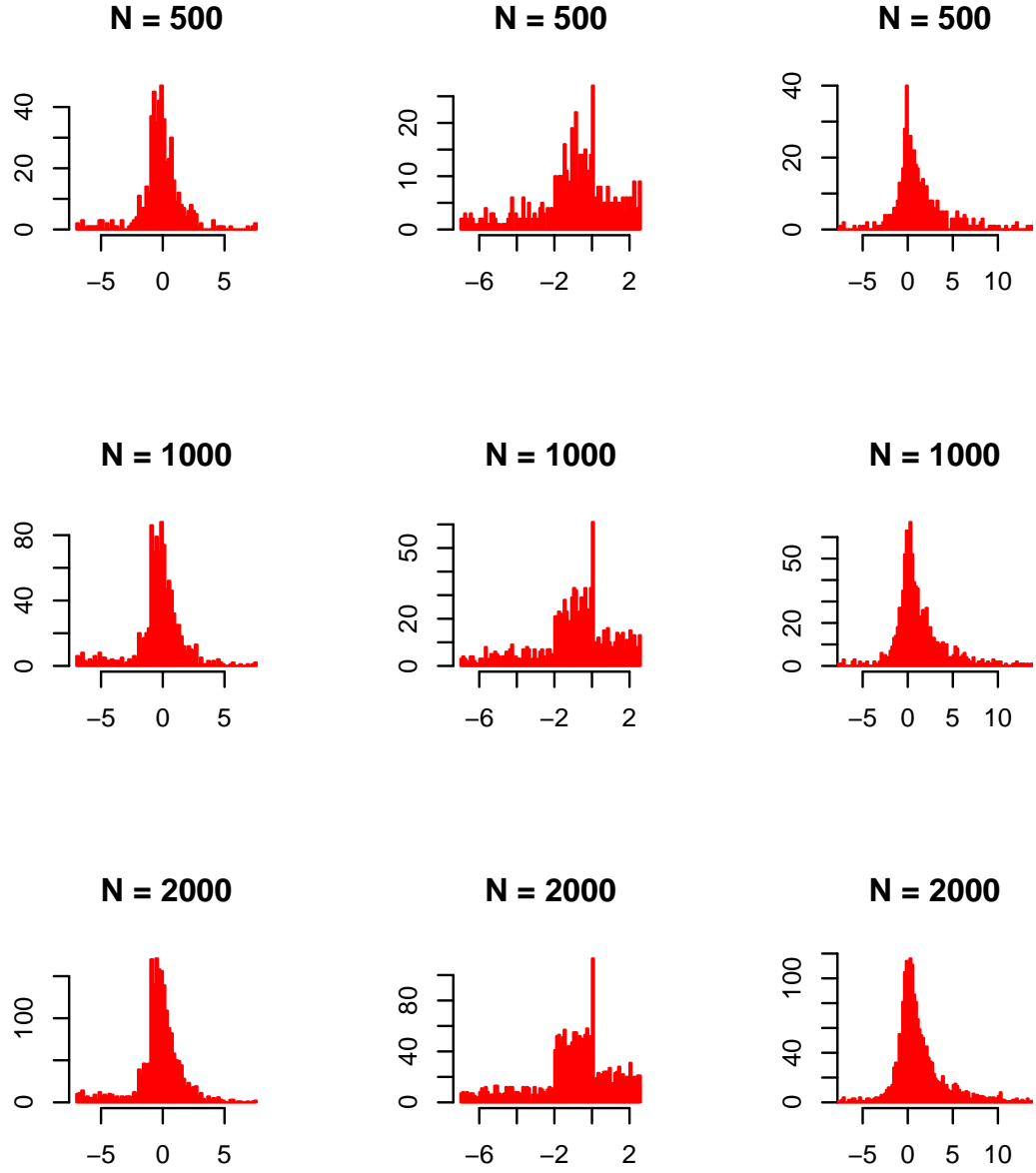


Figure 8: Histograms of the data simulated from unimodal densities  $\lambda_1(t)$  (left column),  $\lambda_2(t)$  (middle column), and  $\lambda_3(t)$  (right column).